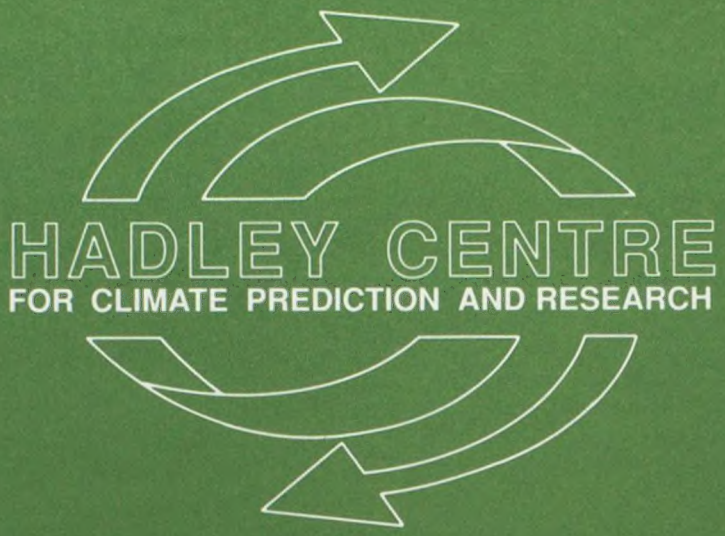

Version 2.2 of the Global sea-Ice and Sea Surface Temperature data set, 1903-1994

by

N A Rayner, E B Horton, D E Parker, C K Folland and R B Hackett

CRTN 74

September 1996

CLIMATE RESEARCH TECHNICAL NOTE

Hadley Centre
Meteorological Office
London Road
Bracknell
Berkshire RG12 2SY

CLIMATE RESEARCH TECHNICAL NOTE NO. 74

VERSION 2.2 OF THE GLOBAL SEA-ICE AND SEA SURFACE TEMPERATURE DATA SET, 1903-1994

by

N A Rayner, E B Horton, D E Parker, C K Folland and R B Hackett

Hadley Centre for Climate Prediction and Research
Meteorological Office
London Road
Bracknell
Berkshire RG12 2SY

NOTE: This paper has not been published. Permission to quote from it should be obtained from the Director of the Hadley Centre.

Version 2.2 of the Global sea-Ice and Sea Surface Temperature data set, 1903–1994.

N. A. Rayner, E. B. Horton, D. E. Parker, C. K. Folland and R. B. Hackett

Abstract

This note presents version 2.2 of the monthly Global sea-Ice and Sea Surface Temperature (GISST) data set. GISST2.2 provides monthly SST and sea-ice fields for the period 1903–1994. GISST2.2 is intended as a forcing data set for Atmospheric GCMs and supersedes the first version, GISST1.1.

We describe the development of the Empirical Orthogonal Function (EOF) technique used in the creation of global fields from incomplete observed data coverage. A coarse global trend EOF and finer 2° latitude by 2° longitude ocean basin EOFs are used to reconstruct the monthly fields for 1949–1981. The fields for 1903–1948 are interpolated using ocean basin EOFs only on an equal area resolution equivalent to 10° latitude by 12° longitude at the equator. Fields for 1982–1994 include the influence of bias-corrected satellite-based Advanced Very High Resolution Radiometer (AVHRR) data and require very little interpolation. A new method of assigning SST to the marginal ice-zones is presented in detail.

GISST2.2 is a significant improvement on GISST1.1. It incorporates a more accurate 1° latitude by 1° longitude 1961–1990 climatology and observed sea-ice concentration data, as well as the improved resolution and analysis methods of the SST fields.

1 Introduction.

Several long historical sea surface temperature (SST) records have been compiled for climate studies (eg. the Meteorological Office Historical SST (MOHSST) data set, 1856 to the present (Parker *et al* (1995b)) and the SST data contained within the Comprehensive Ocean-Atmosphere Data Set (COADS), 1854 to the present (Woodruff *et al* (1987))), but these data sets have considerably less than global coverage because of many data-void areas. The gridded area-averaged values contained within them are largely a collection of manual point samples from ships. Latterly, these "ship observations" have been made in the vessels' water intake pipes and augmented by the automatic measurements of drifting buoys.

These data sets, however, whilst good for empirical climate studies if proper account is taken of the missing data, are insufficient for use as boundary forcing in Atmospheric General Circulation Model (AGCM) experiments. Here, complete global coverage of SST together with sea-ice extent is needed. The first versions of the Global sea-Ice and Sea Surface Temperature data set (GISST1.0 and GISST1.1, Parker *et al* (1995a)), created in the Hadley Centre, were completed and released in early 1993. These monthly mean fields cover the period since October 1948 (although experimental fields back to 1871 were also produced). GISST1.0 and 1.1 have been used by a number of modelling groups to simulate in AGCMs the changes observed in the real atmospheric circulation in recent decades. Such experiments are commonly called Climate of the Twentieth Century integrations; some results of these are included in Folland and Rowell (1995), Gates *et al* (1996), Potts *et al* (1996) and CLIVAR (1995), for example.

Monthly mean SST values in **all versions** of GISST have a nominal 1° latitude by 1° longitude resolution, as the background climatology has been developed on this scale. However, in GISST1.0 and 1.1, the analysis of **anomalies** from the 1951–1980 climatology was carried out on a 5° by 5° spatial resolution (Parker *et al* (1995a)) and the final anomaly resolution is actually slightly coarser because of light smoothing. Collaborative studies with colleagues from the NOAA Climate Prediction Center (eg. comparing with the Reynolds and Smith (1994) optimally interpolated (OI) SST data) have shown that this 5° by 5° resolution is too coarse to adequately represent some recent La Niña events, which produce a narrow band of concentrated cold SST anomalies along the Equator in the East Pacific. Hence, the latest version, GISST2.2, analyses monthly mean SST anomalies for 1949–1994 on a 2° by 2° resolution, the final anomaly resolution again being slightly coarser due to smoothing. GISST2.2 has been extended back to 1903 with analyses of SST anomalies on a near 5° by 5° resolution from the GISST1.2 data set (introduced briefly in Rayner *et al* (1995b) and Rowell *et al* (1996)). The table in Appendix A summarises the various versions of the GISST data set created to date.

Another problem that we have found with GISST1.0, besides its anomaly resolution, is that its methods of filling data-void regions (spatial "growth", weighted temporal interpolation and the Poisson equation technique of Reynolds (1988), see Parker *et al* (1995a) for details) relax interpolated anomalies towards the local 1951–1980 climatology. [1] Hence, we have devised for GISST2.2 a technique which utilises Empirical Orthogonal Functions (EOFs) of SST anomaly variation to interpolate (in 1903–1948) or reconstruct (in 1949–1981) the observed data in a reasonably unbiased way (see Section 4 and Appendix C).

[1] The monthly fields for the period 1982–1994 of GISST1.1, however, are based on much more complete observed data, due to the inclusion of bias-adjusted satellite instrument readings from the Advanced Very High Resolution Radiometer (AVHRR), so this effect is small.

This technique is similar in principle to that used in Smith *et al* (1996). However, the details of the technique differ in the methods of fitting the EOFs to the data and in the length of time series of data used to calculate the EOFs. In addition, the areas over which the EOFs are calculated and the methods used to deal with the trend component of the data are different. Earlier work by Glowienka-Hense and Hense (1986) using gridded Atlantic SST also addressed the problem of filling data-sparse fields using EOFs, albeit in a different manner from both the method presented here and in Smith *et al* (1996).

A further advantage of GISST2.2 over GISST1.1, from 1949 onwards, is the use of a new SST climatology. This was calculated using *in situ* data from the well-observed 1961–1990 period, fitted to a background field including bias-adjusted satellite "observations" (Parker *et al* (1995c)). In particular, it is more accurate in the Arctic and Southern Ocean where some serious local biases existed in the 1951–1980 Bottomley *et al* (1990) climatology on which GISST1.0 and 1.1 were based. That climatology did not have the benefit of remotely sensed data. Indeed, MOHSST6, the *in situ* data set on which the climatology and post-1948 GISST2.2 are based, is itself a significant improvement on MOHSST5 (on which GISST 1.0, 1.1 and 1.2 were based, see Appendix A). MOHSST6 includes newly digitised data and benefits from an improved quality control procedure which utilises a daily version of the 1961–1990 climatology. If the data are considered on a 5° by 5° resolution, there are typically 2% more grid-boxes containing observed data in MOHSST6 than in MOHSST5 (Parker *et al* (1995b)) [2].

GISST2.2, from 1949 onwards, and the background climatology also make use of observed sea-ice concentration data for the Arctic (see Section 2) and new statistically-derived SSTs in the regions where sea-ice and water co-exist. Some erroneous Antarctic sea-ice extent data, used in GISST1.1, have been replaced by observed concentration data. Where sea-ice concentration values are not available from observations, a simple method is used to create them from known ice-edge positions. Having sea-ice concentrations for all parts of the Northern and Southern Hemispheres where sea-ice is found allowed us to derive SSTs wherever there was partial ice coverage (see Section 3).

This note expands upon and updates Rayner *et al* (1995a), detailing the data and methods used to create GISST2.2 and comparing it to the earlier GISST1.1. Section 2 discusses the sea-ice fields. Section 3 introduces the method used to specify ice-zone SST with further details in Appendix B. The development of the open-ocean SST analysis is presented in Section 4, with theoretical details in Appendix C. A discussion of the characteristics of GISST2.2 and the improvements made over GISST1.1 can be found in Section 5. Some thoughts for the future appear in Section 6.

2 Sea-ice concentration and extent data.

Correct representation of ice-edge advance and retreat is vital in a data set used to force AGCMs. For GISST2.2, a new set of gridded 1° latitude by 1° longitude observed monthly Arctic sea-ice concentration data for 1901–1990 has been used. This was obtained on a 60 mile by 60 mile resolution by courtesy of John Walsh of the University of Illinois. A summary of this data set can be found in Walsh (1995). Walsh (1978) and Kelly (1979) describe some of the original data which have been incorporated. Only those data from 1949 onwards were used in GISST2.2, however, as the analysis for 1903–1948 used the

[2]The use of the 1961–1990 climatology is consistent with land air temperatures based on the analysis of Jones (1994), which are also referenced to a 1961–1990 climatology.

GISST1.1 sea-ice extents (see Parker *et al* (1995a) for the origins of these data). The ice-edge in all Arctic sea areas after 1990 was also derived from the GISST1.1 data. With the exception of 1984–1988, the Antarctic sea-ice data are the same throughout the 1903–1994 period as those used in GISST1.1: various climatologies up to 1972 and NOAA observations since 1973 (Parker *et al* (1995a)). For 1984–1988, new concentration data **were** available from the NOAA-Navy Joint Ice Center and replaced erroneous ice-edge position data for this period used in GISST1.1. Sea-ice data for the Caspian Sea, Sea of Japan and Sea of Okhotsk were taken from GISST1.1 ice-edges in all years, as these areas are not covered by the Walsh analysis. The Great Lakes ice concentration data for 1949–1994 were from a NOAA analysis (Assel *et al* (1983)). NOAA data are available in this region only for the period 1960–1979. For 1949–1959 and 1980–1994, a monthly climatology of the 1960–1979 data was used. As in all other areas, the Great Lakes analysis for 1903–1948 used the GISST1.1 sea-ice extents.

The observed Arctic concentrations for 1949–1990 have enabled us to develop a statistical method to derive SSTs for the whole ice-zone (see Section 3). As the new Arctic data do not cover the period 1991–1994, we have derived concentrations for 1991–1994 from ice-edge positions in GISST1.1. We also used GISST1.1 ice-edge positions to derive concentrations for 1949–1994 in all other regions listed above where observed concentration data were not available. Sea-ice concentrations were derived from ice-edge positions as follows:

- Using GISST1.1 ice-edge positions, 1° latitude by 1° longitude areas designated as purely sea-ice or purely sea water were assigned ice concentrations of 10/10 or 0/10, respectively.

- For each 1° by 1° area, the average was calculated of the concentration assigned to it and those of the 8 surrounding 1° by 1° areas.

- Once this was complete, the original concentration assigned to the box was replaced by the calculated average.

This process resulted in smooth gradients of ice concentration. The effect is to convert apparent ice concentration values of 10/10 at the GISST1.1 sea-ice edge to partial ice (see Figure 1). We believe the GISST1.1 ice-edge is best interpreted in this way. The ice-edge in GISST2.2 occurs where the observed or derived concentrations become 10/10. This means, however, that there appears to be a sudden retreat of the ice-edge at the transition between 1948 and 1949. The 1903–1948 analysis used the GISST1.1 sea-ice edge without derivation of a concentration gradient and the 1949–1994 analysis used the new position at 10/10 observed or derived ice concentration.

A file of sea-ice concentrations, along with data-source flags, forms a companion file to the GISST2.2 SST file. The GISST2.2 **SST file** itself holds land and "10/10 sea-ice" masks, denoted by -1.0×10^{30} and -1000 respectively, as well as SST values. Prior to 1949, as the observed and derived ice concentration data were not used, the **ice concentration file** contains 10/10 and 0/10 corresponding to where there was ice or no-ice. After 1949, this file holds the observed concentrations for 1949–1990 in the Arctic and 1984–1988 in the Antarctic and the derived concentrations when and where these were not observed.

3 Ice-zone SST Analysis

SSTs in data-void boxes next to sea-ice were assigned in the 1903–1948 fields of GISST2.2 as in GISST1.1. This was done in three ways (see Parker *et al* (1995a) for details):

- If a climatologically [3] ice-covered 1° latitude by 1° longitude square was open water in any month, it was filled by the average climatological SST of any 1° by 1° open-water squares next to the climatological ice-edge occurring in a 19° by 19° area surrounding the target square.
- Any open-water square adjacent to anomalous ice-covered or anomalous open-water squares was set to its climatological value plus one half of the temperature anomaly in the adjacent anomalous square.
- An open-water square adjacent to a climatologically and presently ice-covered square was set to its climatological value.

In the 1949–1994 analysis, observed and derived sea-ice concentration data in the Northern Hemisphere and the Antarctic (see Section 2) have been used to statistically estimate GISST2.2 SSTs occurring within the marginal ice-zones. The method is introduced briefly in this section; further details can be found in Appendix B.

Observed monthly mean SSTs from *in situ* sources for 1961–1990 in the Northern Hemisphere (1960–1979 in the Great Lakes) and from the satellite AVHRR instrument for 1982–1994 over the Antarctic were regressed against sea-ice concentrations in areas where both SST and sea-ice were found. The resulting equations were used to specify SSTs in all areas with sea-ice concentrations of greater than 0/10 and less than 10/10.

Northern Hemisphere data were treated in the form of (nominally) 360 running 31° longitude bands for each separate 3 month season (Dec–Feb, Mar–May, etc). Each monthly 1° by 1° SST superob was compared to its surrounding 5° by 5° average sea-ice concentration. All such pairs of SST and sea-ice concentration data in a given 31° band in a given season were used to derive a regression relation of the form:

$$SST = aI^2 + bI + c$$

so long as there were at least 100 pairs of data. This was constrained so that when $I = 1.0$ (10/10 ice cover), $SST = -1.8$°C (or 0.0°C in the fresh water Great Lakes which were treated separately). In all 1° by 1° areas with a sea-ice concentration value, the appropriate regression relation was used to calculate an SST value for that box. Figure 2 shows the result of this for a section of the coast of Greenland. Wherever there were too few regional data to form a relation, a longitude-invariant relation using all Northern Hemisphere data (except the Great Lakes) for the appropriate season was used.

NOAA sea-ice concentrations in the Great Lakes were available for the period 1960–1979, enabling us to directly assign ice-zone SST then. At other times, a monthly ice-zone SST climatology calculated over 1960–1979 was used.

For the geographically-simpler Antarctic, only one regression relation was used for all longitude bands in each season. Note that Antarctic *in situ* data were too few to reliably construct these regressions, so satellite data were used. For the Northern Hemisphere, we preferred *in situ* data because these are probably less prone to persistent bias; satellite biases in the ice margin zone are more difficult to correct than in the open ocean.

[3] The climatology period here is 1951–1980.

4 Oceanic SST Analysis

4.1 1903–1948

The monthly fields for the 1903–1948 period of GISST2.2 are taken from an analysis known as GISST1.2, introduced in Rayner *et al* (1995b) and Rowell *et al* (1996). Referring to Appendix A, the reader will note that this was the first attempt to improve upon GISST1.1 using EOF interpolation, according to the theory detailed in Appendix C. It builds upon the same initial observed SST data as GISST1.1 (MOHSST5, corrected for instrumental biases before 1942 as in Folland and Parker (1995)), and is analysed on the same resolution (5° latitude by 5° longitude).

The initial stage of the GISST1.1 interpolation process, the so-called "data growth", was also applied in this analysis, albeit for two rather than three iterations. See Parker *et al* (1995a) for a full account of this procedure. After data-growth, the majority of the remaining gaps were filled using EOFs. Basically, the EOFs were used to fill data-voids by combining the available "grown" *in situ* data and the knowledge of the connections between the data time series at each point of each ocean contained in the EOFs. Missing SST anomalies were built up from a weighted linear sum of EOF patterns, with adjusted time coefficients as weights. These time coefficients were first calculated by projecting the EOF patterns onto the gridded observed data anomalies then adjusted to allow for the likely contribution to this calculation from the missing values. Appendix C explains this process in full with appropriate equations.

The EOFs used for this process were of a coarse resolution version of GISST1.1 anomalies with respect to its 1951–1980 climatology. The analysis was carried out on an equal area grid. The area of each box was equal to that of a box 10° latitude by 12° longitude positioned between the equator and 10°N or S. The use of this coarse resolution aimed to remove the influence of minor irregularities and poor interpolation in some regions of GISST1.1 from the GISST1.2 fields, whilst retaining the essential covariance structure. The monthly GISST1.1 anomaly fields for 1901–1990 were averaged into seasons (Jan.–Mar., Apr.–Jun., etc.) and all seasons together were subject to a covariance EOF analysis. Before this latter step, however, the data were partitioned between each of the three major ocean areas: the Atlantic, Pacific and Indian. Thus three sets of EOFs were obtained.

"Fixed-grid" experiments in each ocean were used to determine the number of EOF modes most skillful in filling data-void areas in the "grown" MOHSST5 data set. These experiments involved reducing the data coverage of 5° by 5° GISST1.1 anomalies for 1981–1990 to the level of some data-sparse period in the MOHSST5 record: 1871–1880 and 1904–1913 were chosen. The data anomalies for each month of 1981–1990 were masked where data for the corresponding month of 1871–1880 or 1904–1913 were missing. These depleted data were then filled using increasing numbers of EOFs. The skill of each interpolation attempt was assessed using the spatial correlation and root mean squared error (RMSE) between the original 1981–1990 GISST1.1 anomalies and the interpolation into the boxes where the values had been suppressed. To overcome the problems of having EOFs defined on one grid and the anomalies gridded onto another, the EOF spatial weights were interpolated from the equal area to the 5° by 5° grid, then renormalised. So that the data being interpolated were independent from the data used in the EOFs, the analysis period of the EOFs used in these experiments was changed to 1901–1980. EOFs for 1901–1980 were very similar to those for 1901–1990, and it was felt that the results of the fixed-grid experiments would be applicable to the latter set. These investigations led

us to use 9 Pacific, 10 Atlantic and 11 Indian EOF modes in the interpolation. At the boundaries between the oceans, the interpolated data were smoothed 1:2:1 east-west.

It was not possible to use these oceanic EOFs to fill gaps over the whole globe, because in some regions there was insufficient observed content in GISST1.1 to be certain that the EOFs were based on observations rather than biased interpolation. For these areas, values were estimated using the Poisson equation technique of Reynolds (1988), with the 1° by 1° resolution 1951–1980 climatology as background field, as in GISST1.1. Assumptions about SST near the ice-edge were as described in Section 3 and Parker *et al* (1995a). This interpolation method leads to some bias towards climatology in some parts of the Southern Ocean.

4.2 1949–1981

The EOF interpolation technique used to complete the monthly fields for 1903–1948 (Section 4.1) was substantially modified for the analysis of the 1949–1981 fields. A key difference is that in this period the original *in situ* observed data are not filled by the EOF reconstructed anomaly fields, but replaced altogether, as in Smith *et al* (1996).

Like the fields for 1903–1948, the 1949–1981 fields are based on a version of MOHSST. However, the latter fields build upon the latest version, MOHSST6 (Parker *et al* (1995b)) which has the advantage of extra data and an improved method of quality control, utilising for this a daily version of the new 1961–1990 climatology (Parker *et al* (1995c)). Like MOHSST5, MOHSST6 contains added COADS data (Woodruff *et al* (1987)). We have used a version of MOHSST6 on 2° latitude by 2° longitude resolution, both for the base of the final analysis and for the calculation of the EOFs.

Using a very long data series to calculate the EOFs (as in Section 4.1) ensures that the global trend is captured within them, but the use of a very coarse resolution raises questions about the reliability of the analysis on the local scale. Thus, it was decided to split the analysis in two, using a coarse resolution global EOF calculated from a long record to capture the trend and sets of finer resolution ocean basin EOFs of a shorter, better observed period to provide the local details.

The global trend was isolated very well using the first EOF of a low-pass filtered version of GISST1.1 anomalies (A. W. Colman personal communication). The seasonal GISST1.1 anomalies were on the same coarse equal area grid as used in Section 4.1. They were filtered using a Kalman filter which retained variability of period greater than 8 years. Care was taken to filter a longer period than actually required so as not to lose any information in the filtering process. An EOF analysis of these data using all filtered seasons for 1901–1990, has a first mode with the spatial pattern and time series shown in Figure 3. It is quite similar to the first EOF of SST anomaly on 10° latitude by 20° longitude resolution shown by Parker and Folland (1991). Our global coarse resolution EOF1 is later used to reconstruct the trend component of the monthly fields, but first it was used to de-trend the data used to calculate the detailed oceanic EOFs.

Seasonal anomalies for 1951–1990 of the 2° by 2° resolution MOHSST6 were first filled using a background field, as a spatially and temporally complete set of fields is required for any EOF analysis. The background field was reconstructed using 20 global-scale coarse equal area EOFs of unfiltered GISST1.1 anomalies for 1901–1990. This coarse reconstruction was re-gridded into the 2° boxes and used to fill the gaps in the MOHSST6 field. The global trend component was then removed using the first EOF of the low-pass filtered data. These filled, de-trended, 2° resolution data were then divided into four

regions: the Atlantic Ocean, Indian Ocean, Pacific Ocean and the Mediterranean/Black Sea region. An all seasons area-weighted covariance EOF analysis was carried out on each of the four regions. Care was taken to exclude any grid boxes where an insufficient amount of observed data made up the series. Also, the ocean areas were designed to overlap slightly to allow some smoothing of the reconstructed fields at the boundaries.

Figures 4–7 show the leading four of these EOFs in each region. As we are really only interested in capturing and thus producing a good representation of the variance of the SST anomaly data, we do not look for recognisable patterns in these EOFs. Nevertheless, especially in the Pacific, we are able to see some well-known phenomena related to ENSO. Figure 8 is a plot of natural log of eigenvalue against mode number for the first 50 Pacific modes. This was used to determine how the modes were grouped. The modes which have been marked on the plot all occur after a "shelving" of the curve. Where such a shelf occurs, the eigenvalues of two modes are very similar; this is an indication that the modes associated with these eigenvalues are mixed. Modes beyond the last of these visible shelves are likely to be noise if the plot becomes linear. This can be seen on Figure 8 after EOF 30.

The theory used to produce reconstructions of whole fields of anomalies is the same as used in Section 4.1 and detailed in Appendix C. However, as the period under reconstruction is one of good observed data coverage and for the most part not independent of that of the data from which the EOFs were calculated, we felt that "fixed-grid" experiments of the kind used to determine the most skillful number of EOFs to interpolate the 1903–1948 period would not be necessary. Instead, we tried a selection of numbers of modes chosen using plots of natural log of eigenvalue versus mode number, as described above. Figure 8 for the Pacific has marked on it those modes tested for that ocean. Table 1 shows two measures of the skill of reconstructions of 1982–1993 fields using these four sets of Pacific modes. After reconstructing using each set of de-trended Pacific modes and adding the contribution reconstructed from the global trend EOF, the reconstructions were smoothed 1:2:1 east-west then north-south. The skill was assessed on a monthly and seasonal basis. The reconstructions were compared on 2° by 2° resolution with both monthly and seasonal OI SST (Reynolds and Smith (1994)) and independently created 1982–1993 GISST2.2 (Section 4.3) anomalies with respect to the same 1961–1990 climatology.

no. of EOFs used	OI				GISST2.1			
	monthly		seasonal		monthly		seasonal	
	corrl.	rmse (°C)	corrl.	rmse (°C)	corrl.	rmse (°C)	corrl.	rmse (°C)
4	.48	.62	.52	.54	.50	.53	.58	.41
10	.55	.59	.60	.51	.59	.49	.67	.38
21	.61	.57	.65	.49	.66	.46	.74	.35
30	.63	.56	.67	.48	.69	.44	.76	.33

Table 1: Time averaged statistics of Pacific reconstructions using varying numbers of oceanic EOFs and the global trend EOF. Compared with both OI SST and GISST2.1.

Although the improvements are fairly small, the skill appears to continue to rise until 30 EOFs were used. Figure 9 compares the seasonal variance of the OI SST anomalies and reconstructions using 4, 21 and 30 EOFs. These plots lend weight to the argument that 30 EOFs was the best number to use in the Pacific. We can see, for example, from Figure 9(c) and (d) that using 30 rather than 21 EOFs increases the coherence of the variance

of the equatorial East Pacific.

After going through the same process in all four EOF areas, we decided to use 32 EOFs in the Atlantic and Indian Oceans, 30 in the Pacific Ocean and 15 in the Mediterranean/Black Sea region. The oceanic anomaly fields reconstructed from these modes were subject to a spacewise linearly-varying weighted average across the overlapping parts of the EOF domains, then the influence of low frequency EOF1 was added back.

Like the EOFs used to interpolate the 1903–1948 fields, those used to reconstruct the 1949–1981 fields are not truly global in their coverage: for example, there were too few data in the Southern Ocean and parts of the south-eastern Pacific to include in the EOFs. In some regions, outside of the range of influence of the EOFs and the marginal ice-zone and where the data are particularly sparse or unreliable, it was necessary to use fixed monthly climatologies in place of annually varying data. In the Caspian and Aral Seas, monthly climatologies of AVHRR "observations" for 1982–1994 were used in all years. *in situ* observations were inserted, where available, into the Great Lakes, Hudson and Baffin Bays and within the Canadian Archipelago. These were in many cases superseded by the regressed ice-zone SST (Appendix B). The observations and ice-zone SSTs anchored the Poisson equation technique which was used to fill the remaining data-void areas (as in Section 4.1) taking the 1° by 1° 1961–1990 climatology as background field. Anomalies of these completed fields were then lightly smoothed on 2° by 2° resolution (1:2:1 east-west then north-south), and re-converted to SST values with a 1° by 1° resolution by adding back the 1961–1990 climatology.

4.3 1982–1994

From 1982 onwards, GISST2.2 is similar in construction to GISST1.1 (Parker *et al* (1995a)) in that both make use of bias-corrected satellite SST data from the AVHRR instruments and EOF reconstruction is not used. However, as already explained in this note, GISST2.2 analyses anomalies on a 2° latitude by 2° longitude resolution, compared to 5° by 5° in GISST1.1.

Complete SST fields were produced in two simple steps:

- Uncorrected 2° by 2° satellite SST anomalies were added onto a set of 1° by 1° 1961–1990 normals and collated with ice-zone SSTs. These monthly fields were then filled using the Poisson equation technique with the 1° by 1° climatology as background field.

- Fields of 2° by 2° *in situ* MOHSST6 anomalies were then filled using the same technique with the output of the first step as the background field and the ice-zone SSTs as ice-edge boundary conditions.

This process effectively corrects biases in the satellite data, as the satellite observations themselves are not directly used. Monthly fields were converted to 2° by 2° resolution anomalies from the 1961–1990 climatology and lightly smoothed (1:2:1 east-west then north-south), then re-converted to SST values on a 1° by 1° resolution by adding back the 1° by 1° 1961–1990 climatology.

In the Aral Sea, as for 1949–1981, a monthly climatology of 1982–1994 AVHRR "observations" was used.

5 Properties of GISST2.2

In this section, we will examine the properties of the GISST2.2 data set and compare it with GISST1.1. We will start by discussing the effect that developing the data set in three separate temporal pieces has had on its continuity, then pick the Black Sea, an area of high variability, for closer scrutiny. Finally, we will show the improvements to be found in the representation of Pacific warm and cold events.

5.1 Discontinuities in 1948/1949

Recalling from earlier sections that the monthly fields for 1903–1948 were taken from the GISST1.2 analysis, which uses the GISST1.1 ice-edge, a different way of specifying ice-zone SST, and a coarser anomaly resolution than used in the 1949–1994 analysis, the reader will not be surprised that there are associated discontinuities in the data set.

As mentioned in Section 2 there will appear to be a sudden retreat of the ice-edge in 1949 due to the change in the way that the ice-edge is treated. Also, Arctic ice-zone SSTs become warmer after 1948, as the method used to assign them in GISST1.1 and 1.2 was cold-biased.

In some parts of the open ocean and inland sea regions there are also small step changes in the SST in 1949. The changes seem generally to be around 0.5°C, with local changes of up to 1°C in places. Many of these discontinuities in the data-poor regions will be due to the different background climatologies used (1951–1980 means were used prior to 1949 and 1961–1990 means were used from 1949) as the Poisson equation technique used in especially data-poor areas tends to relax the interpolated SST towards this.

5.2 The Black Sea

There are three very anomalous periods in the 1949–1994 Black Sea data. October 1951 was reconstructed using information from Mediterranean data only as none were available in the Black Sea. However, the large negative anomalies reconstructed in this month are supported by the large negative anomalies in that region in the land/sea (CRU/MOHSST6, Jones (1994) and Parker *et al* (1995b)) temperature data set. February 1956 is a similar case, which is supported by very anomalous cold temperatures over much of Europe. The very large negative anomaly which can be seen in the Black Sea average for June 1976 is due entirely to the observed data in MOHSST6 there.

5.3 Comparing GISST2.2 with GISST1.1

Although the Niño regions' [4] average time series (not shown) are little different between GISST2.2 and GISST1.1, improvements can be seen in the spatial representation of single El Niño and La Niña episodes. Figures 10 and 11 show two well known examples of such events. The La Niña in progress in May 1988 is shown in Figure 10; the two areas of negative anomaly separated by postive anomaly at around 120°W in GISST1.1 have become a single negative area spanning 160°E to the coast of Equador and Peru in GISST2.2. The anomalies associated with this event also have greater amplitude in the new data set. The El Niño event which reached its peak in January 1983, shown in

[4] ie. the regions defined by NOAA/CPC for climate monitoring (used also in Barnston and Ropelewski (1992)): Niño 1 (90°–80°W, 5°S–10°S), Niño 2 (90°–80°W, 0°–5°S), Niño 3 (5°N–5°S, 150°–90°W) and Niño 4 (5°N–5°S, 160°E–150°W).

Figure 11, is also much more coherent in the new version of GISST. Interestingly, the most positive area in Figure 11(a) has been lost in 11(b). The strength of this anomaly over several degrees of latitude and longitude in Figure 11(a) appears on further investigation to have been largely a product of the 5° by 5° resolution of the MOHSST5 data from which GISST1.1 was formed. At this resolution, in this relatively data-sparse region, *in situ* observations can be attributed to much larger areas than may be valid. Less was made of this very small patch of large anomalies in the finer 2° by 2° MOHSST6 and so it has not been preserved through the satellite blending and smoothing stages. The finer resolution of the GISST2.2 analysis between 1949 and 1994, along with the improved sea-ice data, new ice-zone SST specification and more accurate background climatology make for much more accurate global SST and sea-ice fields than were available from GISST1.1.

6 Discussion

This note has detailed the development of the latest release of the GISST data set, GISST2.2. It is, without doubt, an improvement on GISST1.1 for many reasons. The monthly fields for 1949–1994 in GISST2.2 contain more accurate sea-ice data, a better representation of near-ice SST, a more representative background climatology and a higher resolution SST analysis. This section, however, will discuss some parts of the analysis which could be improved upon and introduce some thoughts for future releases.

Data for the period 1903–1948 were not released in GISST1.1, so the extension of GISST2.2 back to 1903 is an improvement in itself. However, many parts of the analysis of these data in particular need further development. As mentioned in Section 4.1 the analysis resolution, 5° by 5°, is the same as that of GISST1.1. This is definitely too coarse to produce optimal analyses. As a result of this the strong La Niña event of 1917, seen in the Southern Oscillation Index (not shown), does not appear. Also 1903–1948 contains the periods of the two World Wars, which have left a legacy of very poor observed SST data coverage. The combination of coarse resolution, sparse data and the method of "data growth", whereby anomalies are spread to their data-void neighbours, is likely to have increased the influence of unrepresentative boxes. These problems cannot be reversed by EOF interpolation into data-void boxes alone.

A further version of GISST, version 2.3, is now under development. This extends the fields back still further to 1871 and reanalyses the fields up to 1948. This analysis includes all the new Walsh sea-ice, the improved ice-margin SST specification, and the climatology used in the post-1948 part of GISST2.2 and reconstructs SST anomaly fields in a similar way. EOFs used in GISST2.3 are a global trend EOF and 4° by 4° resolution ocean basin EOFs. The "data growing" step is left out, allowing the EOF reconstruction to be based on truly representative observations only. The anomaly analysis is done on a 4° by 4° grid which allows a grid box to span the equator. This leads to considerably better representation of La Niña events. The periods of sparse data coverage simulated in fixed-grid experiments (see Section 4.1) have been varied in the development of GISST2.3. This has resulted in different numbers of EOFs being used according to the available observed data coverage. GISST2.3 will be the subject of a forthcoming paper.

We consulted colleagues at NCEP during our deliberations on whether to interpolate or reconstruct the observed fields for 1949–1981. The balance of opinion then was that we should reconstruct the fields completely as this would provide a smoother input for AGCMs. However, this is a question which has still not been completely resolved. The main objection to inclusion of the observations is that it introduces inhomogeneities and

discontinuities between the reconstruction and the observations. We plan to overcome this ultimately, by producing a version of the data set formed by combining the reconstruction and observations using the Optimum Interpolation technique. In the meantime, a version of GISST2.2 is in preparation which includes the MOHSST6 *in situ* observations using a data-adaptive smoothing routine to minimise the inhomogeneities. This will be documented fully in a later text. GISST is a continually evolving data set.

GISST2.2, as described in this note, is being used as the SST forcing for the period before the availablilty of satellite data in the NCEP Reanalysis project (Kalnay *et al* (1996)) and is already being used in several AGCM integrations internationally.

Acknowledgements

We are indebted to Dr. Neil Ward for his advice throughout the course of this project. Thanks go to our colleagues at NCEP: Tom Smith, Bob Livezey and Dick Reynolds for their input. Nick Rayner also wishes to thank David Sexton for his programming advice and useful discussions.

A Versions of GISST

Version	Description	Period
1.0	Anomalies analysed on 5° by 5° spatial resolution. Data "growth" and weighted temporal interpolation. Based on 1951–80 climatology. Sea-ice extent data only. Ref: Parker *et al* (1995a).	Oct 1948–Dec 1981 (with experimental data back to 1871)
1.1	Anomalies analysed on 5° by 5° spatial resolution. Based on 1951–80 climatology. Sea-ice extent data only. Incorporates influence of satellite data. GISST1.0 and 1.1, joined are known as GISST1.1. Ref: Parker *et al* (1995a).	Jan 1982–Dec 1994
1.2	Anomalies analysed on 5° by 5° spatial resolution. Data "growth" and spatial interpolation with coarse equal area ocean basin SSTA EOFs. Based on 1951–80 climatology. Sea-ice extent data only. Ref: Rayner *et al* (1995b).	Jan 1903–Dec 1981
2.1	Anomalies analysed on 2° by 2° spatial resolution. Based on 1961–90 climatology. Sea-ice concentration data. Incorporates influence of satellite data. Statistical ice-zone SST.	Jan 1982–Dec 1994
2.2	Anomalies analysed on 2° by 2° spatial resolution. Data reconstruction with 2° by 2° ocean basin and coarse equal area global EOFs. Based on 1961–90 climatology. Sea-ice concentration data. Statistical ice-zone SST. When joined with 1.2 and 2.1, the whole period 1903–94 is known as GISST2.2.	Jan 1949–Dec 1981

B Ice-zone SSTs

This appendix supplies details of the derivation of marginal ice-zone sea surface temperatures (SSTs) for the period 1949–1994 in GISST2.2. Some details of Rayner *et al* (1995a) and Parker *et al* (1995c) are superseded. The relationships between SST and sea-ice concentration are presented and briefly discussed.

B.1 Derivation of relationships between sea-ice concentration and SST.

Monthly *in situ* SST superobs for 1961–1990 in Northern Hemisphere 1° by 1° areas with non-zero sea-ice concentration were rounded to the nearest 0.5°C and then related statistically to monthly concentric 5° by 5° area sea-ice concentrations I using quadratic least-squares regressions of the form:

$$SST = aI^2 + bI + c$$

The regressions (eg. Figures 12 and 13(a) and (b)) were calculated after the temperatures had been averaged for each ice-concentration class from 0.1 (1/10 ice cover) to 0.9 (9/10 ice cover), with additional specifications that

- when $I = 1.0$ (10/10, complete ice cover), $SST = -1.8$°C (with one exception given below); and
- that $d(SST)/dI$ was constrained to be negative or zero for all I as illustrated in Figure 12

The sea-ice concentration data were mainly taken from the new analysis by John Walsh (Walsh (1995) and see Section 2), except in the Great Lakes where concentrations were taken from a NOAA analysis (Assel *et al* (1983)). For each season (December to February, etc. not January to March, etc. as stated in Rayner *et al* (1995a) and Parker *et al* (1995c)), separate relationships were developed for overlapping 31° longitude bands 0°–31°E, 1°–32°E, ..., 1°W–30°E to avoid discontinuities in SSTs specified using them. The width of the bands was 31°, not 30° as stated in Rayner *et al* (1995a) and Parker *et al* (1995c), in order for central target 1° by 1° areas to be definable. The relationships were not varied with latitude, except that the freshwater Great Lakes were treated separately, by developing seasonal relationships for 1960–1979 (the period of available ice-concentration data) with the specification that when $I = 1.0$, $SST = 0.0$°C. Relationships were not constructed from fewer than 100 pairs of observations. A set of relationships using data irrespective of longitude, but excluding the Great Lakes, was constructed for use in areas with too few data to define regional relationships.

For the Antarctic, a set of relationships for 1982–1994 was constructed in the same way, using data irrespective of longitude. However the SSTs were measured by the satellite-based Advanced Very High Resolution Radiometer (AVHRR), and the sea-ice concentration data were derived from the GISST1.1 (Parker *et al* (1995a)) ice-edge except for the period 1984–1988 when observed concentrations were available. As detailed in Section 2, a gradual ice-edge was derived for the periods before and after 1984–1988 by smoothing sea-ice concentrations assigned to 1° by 1° ice and sea boxes in the GISST1.1 data set. Ice boxes were initially assigned a concentration of 10/10 and sea boxes one of 0/10. By

replacing these concentrations by the average of themselves and the 8 nearest neighbouring concentrations, a gradual progression of ice-free through partial ice to complete ice was achieved.

B.2 Application of the relationships.

With two exceptions as listed below, SSTs in the Northern Hemisphere were specified in each 1° by 1° area having non-zero sea-ice concentration using the relationship for the longitude band centred on the target 1° by 1° area for the appropriate season and the sea-ice concentration in the same 1° by 1° area in the individual month, eg. the SST in the box centred on 0.5°W 75.5°N for January 1949 was calculated using the Dec.–Feb. relation for the 16°W–15°E longitude band. Although the sea-ice concentration data for the majority of the Northern Hemisphere were from the new Walsh analysis, in some peripheral regions (Caspian Sea, Sea of Japan, Sea of Okhotsk) the concentration data were derived from the GISST1.1 ice-edge (as described in Section 2 and Section B.1). The two exceptions were:

- The Great Lakes were treated separately, using the single set of relationships described in Section B.1 (see Figure 13(c)). Marginal ice-zone temperatures were not specified in the Great Lakes in summer and autumn, because the occurrence of ice was too rare to allow relationships to be constructed. As there were no ice data before 1960, a monthly climatological average of specified ice-zone SST for 1960–1979 was used in the Great Lakes for 1949–1960 and after 1979.

- In the areas (apart from the Great Lakes) and seasons for which relationships were not derived owing to paucity of data, the longitude-invariant relationships described in Section B.1 and pictured in Figure 13(a) were used.

For the Antarctic, the longitude-invariant relationships derived using AVHRR SSTs (see Section B.1 and Figure 13(b)) were used in the same way.

In all areas with sea-ice, calcuated SSTs superseded observed values. This avoided the creation of "bulls'-eyes" by occasional isolated observations.

B.3 Description and discussion of the relationships

Figure 12 shows the relationships for each season for a selection of the overlapping 31° longitude bands of the Northern Hemisphere centred at 70.5°W, 30.5°W and 129.5°E. The mean temperatures for each ice-concentration class are also plotted, along with the total number of pairs of observations. Figure 13(a) and (b) respectively show the Northern Hemisphere and Antarctic longitude-invariant relationships, class-mean temperatures, and numbers of pairs of data. Figure 13(c) presents the corresponding results for the Great Lakes.

We know of no previous large compilation of marginal ice zone SSTs, and draw attention to the following features:

- As expected, for a given sea-ice concentration the ocean surface is warmer in summer than at other seasons. This is evident at most longitudes and is clearly shown in the Northern Hemisphere longitude-invariant analyses in Figure 13(a). The surface waters are warmed by absorption of insolation.

- For a given sea-ice concentration, the ocean surface is warmer in the Iceland-Greenland region than at other longitudes (Figure 12(b)), especially in autumn, winter and spring. In this region, sea-ice is advected equatorwards across a strong SST gradient. The contrast with other regions is reduced in summer when the solar heating of the mixed layer dominates.

- In winter, the Great Lakes have warmer surface waters than the Northern Hemisphere average for a given sea-ice concentration (compare Figures 13(a) and (c)). Temperatures are around 3°C for up to 50% ice-cover, probably as a result of the density maximum of fresh water near 4°C and the mixing of underlying water at this temperature with surface water by strong winds. In the first versions of GISST2.2, Great Lakes data dominated the regressions which were used to calculate winter and spring SSTs in the Hudson Bay, where near-surface salinity is 28 to 33 p.s.u. in winter and 25 to 32 p.s.u. in summer, ie. not much less than the typical oceanic value of 35 p.s.u. (Pickard and Emery (1982)). As a result, the prescribed temperatures for the Hudson Bay were too high. In the revised data set, Hudson Bay surface temperatures are mainly calculated from the longitude-invariant relationships (excluding Great Lakes data) in winter and spring owing to a paucity of local data, but regional relationships, excluding Great Lakes data, are used in summer and autumn (not shown).

- In the Northern Hemisphere, typical ocean surface temperatures for 50% ice cover are 1°C to 1.5°C in winter and spring, 0.5°C in autumn (when heat loss dominates) and 2.5°C to 3°C in summer (see above). As a result GISST2.2 is much warmer in the marginal ice zones than the optimum interpolation (OI) analysis of Reynolds and Smith (1994) who set the SST to -1.8°C in areas with 50% or more ice.

- In both hemispheres, observed SSTs for 90% ice cover systematically exceed the fitted relationships. The appearance of this feature in the AVHRR SSTs in Figure 13(b) shows that it is not a result of ships' sampling of the larger open leads in heavily iced seas. A linear fit in the range 10% to 90% ice cover, followed by linear interpolation to the freezing point at 100% cover, may be a better formulation for future verions of GISST.

- For a given sea-ice concentration, AVHRR SSTs in the Antarctic (Figure 13(b)) are in the range 2–4° lower than in the longitude-invariant Northern Hemisphere analyses (Figure 13(a)). The summer (December to February) shows only slightly higher temperatures than the other seasons for a given sea-ice concentration. Although there are differences in the heat balance of the ocean surface between the Arctic and Antarctic, and ships may yield a warm bias by avoiding the waters closest to ice, the differences in the plots may have been enhanced by biases in the satellite data which are calibrated using buoys on a global, rather than regional, basis. The possibility of volcanically-induced cold biases was tested by repeating the regressions with AVHRR omitting 1982–83 (the period of influence of the eruption of El Chichón) and 1991–92 (Mt. Pinatubo): the new regressions did not differ by more than 0.2°C from the old ones, eliminating volcanoes as a major contributor to the differences bewteen the Northern and Southern Hemisphere regressions. We plan, for future versions of GISST, to recalculate the Antarctic relationships using Along-Track Scanning Radiometer (ATSR) SSTs, when these are available with improved cloud-clearing, ice-clearing and other quality-controls (Jones *et al* (1996)).

C Theory of EOF Reconstruction

A covariance Empirical Orthogonal Function (EOF) analysis provides an alternative description of the behaviour of data time series at m positions on a grid in terms of spatial patterns of numbers on the same grid. These patterns are the orthogonal eigenvectors, $\underline{A}_1, \underline{A}_2, \ldots, \underline{A}_m$, of the covariance matrix of the data. At each point, i, the data time series, $\underline{X}_i$, can be reformed by a weighted sum of these patterns.

$$\underline{X}_i = \sum_{k=1}^{m} A_{ik}\underline{Z}_k = A_{i1}\underline{Z}_1 + A_{i2}\underline{Z}_2 + \ldots + A_{im}\underline{Z}_m \tag{1}$$

where A_{ik} is the loading of pattern $\underline{A}_k$ at position i. The weights, $\underline{Z}_k$, will be referred to as the time coefficients of EOF k. The larger the time coefficient of EOF k at any particular point in time, t, the more the field at that time is governed by pattern k. Negative time coefficients indicate a reversal of sign of the numbers in the pattern.

Assuming that the covariances and general behaviour of the data for the time period from which the EOFs are calculated are applicable to other periods of the same variable (in this case SST anomalies) we can use our spatial patterns to reconstruct earlier, more data-sparse, periods. The following theory is due to Ward (1989).

In order to reconstruct the data time series exactly at each point using Equation 1 we must first calculate the time coefficient of each EOF at time t, Z_{kt}. **If the field at time t were part of the set of complete fields from which the EOFs were calculated** this would be an easy task using:

$$Z_{kt} = \sum_{i=1}^{m} A_{ki}X_{it} \tag{2}$$

But, if some data are deleted from these fields, we cannot calculate each Z_{kt} as simply as in Equation 2, since we do not know all the $X_{1t}, \ldots, X_{mt}$. However, the time coefficient of each eigenvector for each time t, Z_{kt}, may be resolved into two contributions: that from the missing data, Z'_{kt}, and that from the available data Z^*_{kt}:

$$Z_{kt} = Z'_{kt} + Z^*_{kt} \tag{3}$$

We can easily calculate Z^*_{kt} for any particular time t by summing over all available data in that field, in an expression similar to Equation 2:

$$Z^*_{kt} = \sum_{i^*} A_{ki^*}X_{i^*t} \tag{4}$$

(here i^* indicates the positions of the available data). The same could be done for Z'_{kt}, summing over the missing data (if we knew them):

$$Z'_{kt} = \sum_{i'} A_{ki'}X_{i't} \tag{5}$$

(here i' indicates the positions of the missing data). Since we do not know what the missing data are, some estimation is required. Manipulation of these equations leads us to an unbiased estimate.

The **actual**, but unknown, value of the missing observation in a grid box at time t is given by (see Equation 1):

$$X_{i't} = \sum_{k=1}^{m} A_{i'k} Z_{kt} = A_{i'1} Z_{1t} + A_{i'2} Z_{2t} + \ldots + A_{i'm} Z_{mt} \tag{6}$$

This expression for $X_{i't}$ can then be substituted into Equation 5 to give Z'_{kt} in terms of the time coefficient over the whole field, Z_{kt}:

$$\begin{aligned} Z'_{kt} &= \sum_{i'} A_{ki'} X_{i't} \\ &= \sum_{i'} A_{ki'} \sum_{j=1}^{m} A_{i'j} Z_{jt} \end{aligned}$$

substituting for Z_{jt} from Equation 3 gives us

$$\begin{aligned} Z'_{kt} &= \sum_{i'} A_{ki'} \sum_{j=1}^{m} A_{i'j} (Z'_{jt} + Z^*_{jt}) \\ &= \sum_{j=1}^{m} (\sum_{i'} A_{ki'} A_{i'j}) (Z'_{jt} + Z^*_{jt}) \end{aligned}$$

Now, $\sum_{i'} A_{ki'} A_{i'j}$ is a constant at a given point in time for each pair of j and k. We will call this constant C'_{kj}. Calculating all such pairs of j and k, we form a matrix $\underline{C'}$ for each point in time. $\underline{C'}$ varies as the positions of the missing data, i', vary. Each element in this matrix, C'_{kj}, is the sum over all missing data positions of the product of the components of $\underline{A_j}$ and $\underline{A_k}$ at those positions. Thus:

$$Z'_{kt} = \sum_{j=1}^{m} C'_{kj} (Z'_{jt} + Z^*_{jt}) \tag{7}$$

which is true for the time coefficients of all m EOFs, so:

$$\underline{Z'_t} = \underline{C'} (\underline{Z'_t} + \underline{Z^*_t}) \tag{8}$$

where the one-dimensional matrices $\underline{Z'_t}$ and $\underline{Z^*_t}$ contain the unknown and known contributions to the time coefficients of all eigenvectors at time t. Equation 8 provides us with an expression for $\underline{Z'_t}$ in terms of the known $\underline{Z^*_t}$ and $\underline{C'}$:

$$\begin{aligned} (\underline{I} - \underline{C'}) \underline{Z'_t} &= \underline{C'} \underline{Z^*_t} \\ \underline{Z'_t} &= (\underline{I} - \underline{C'})^{-1} \underline{C'} \underline{Z^*_t} \end{aligned} \tag{9}$$

which is valid providing $(\underline{I} - \underline{C'})^{-1}$ exists. Finally, to reassemble the whole field at time t, $\underline{X_t}$, we need the whole time coefficient for all m EOFs at that time, $\underline{Z_t}$ where:

$$\begin{aligned} \underline{Z_t} &= \underline{Z'_t} + \underline{Z^*_t} \\ &= ((\underline{I} - \underline{C'})^{-1} \underline{C'} + \underline{I}) \underline{Z^*_t} \\ &= (\underline{I} - \underline{C'})^{-1} \underline{Z^*_t} \end{aligned} \tag{10}$$

Then

$$\underline{X}_t = \underline{A}\,\underline{Z}_t \tag{11}$$

where $\underline{A}$ is the complete matrix of all EOF patterns.

In practice, the fields we are completing are not likely to be from the set used to calculate the EOFs. Also, the EOFs may contain some small-scale noise, which we do not want in our reconstruction. We cope with this by using only a subset of the EOFs, which is chosen to minimise noise and over-fitting to the available data. It is assumed that the leading modes are more likely to represent variance common to different sets of the same variable. Both the use of only a subset of modes and the use of EOFs calculated from one data set to reconstruct another mean that our method can no longer be called completely unbiased. Careful choice of the EOFs used leads us to a method which is as unbiased as possible.

This method is similar in principle to that of Smith *et al* (1996), but differs in practice. In Smith *et al* (1996) the EOFs are fitted in a least squares sense to the data, which takes the place of the adjustment of the time coefficients by $\underline{C}'$ detailed above. Glowienka-Hense and Hense (1986) took account of the contribution from missing data to the time coefficients in a different manner. They inflated the EOF time coefficent by multiplying by the ratio of the total number of boxes in the grid to the number of boxes which contained observed data. We believe that our method, based on Equation 10, is in principle more exact as we have taken account of the varying "importance" of the grid boxes.

References

Assel, R. A., F. H. Quinn, G. A. Leshkevich, and S. J. Bolsenga, 1983: *Great Lakes Ice Atlas.* Great Lakes Environmental Research Laboratory, Ann Arbor, Michigan 48104.

Barnston, A. G., and C. F. Ropelewski, 1992: Prediction of ENSO Episodes Using Canonical Correlation Analysis. *Journal of Climate*, **5**(11), 1316–1345.

Bottomley, M., C. K. Folland, J. Hsiung, R. E. Newell, and D. E. Parker, 1990: *Global Ocean Surface Temperature Atlas.* HMSO, London. Publication funded by UK depts of Energy and Environment.

CLIVAR, 1995: *CLIVAR. A study of climate variability and predictability. Science plan. World Climate Research Programme. WCRP 89 WMO/TD 690.*

Folland, C. K., and D. E. Parker, 1995: Correction of instrumental biases in historical sea surface temperature data. *Quarterly Journal of the Royal Meteorological Society*, **121**, 319–367.

Folland, C. K., and D. P. Rowell, editors, 1995: *Workshop on simulations of the Climate of the Twentieth Century using GISST, 28–30 November 1994, Hadley Centre, Bracknell, UK.*, Meteorological Office, London Road, Bracknell, Berkshire, RG12 2SY. Hadley Centre for Climate Prediction and Research. CRTN 56.

Gates, W. L., A. Henderson-Sellers, G. J. Boer, C. K. Folland, A. Kitoh, B. J. McAvaney, F. Semazzi, N. Smith, A. J. Weaver, and Q.-C. Zeng, 1996: Climate Models-Evaluation. In Houghton, J. T., L. G. Meira Filho, B. A. Callandar, N. Harris, A. Kattenberg, and K. Maskell, editors, *Climate Change 1995. The Science of Climate Change.*, chapter 5, 235–284. Cambridge University Press, Trumpington Street, Cambridge.

Glowienka-Hense, R., and A. Hense, 1986: Completion of an incomplete data set by EOF-analysis: the method is applicated to the historical sea surface temperature data set. In *Proceedings of the Third International Conference on Statistical Climatology, June 23-27, 1986, Vienna, Austria.*, 22–27.

Jones, M. S., M. A. Saunders, and T. H. Guymer, 1996: Reducing cloud contamination in atsr averaged sea surface temperature data. *Journal of Atmospheric and Oceanic Technology*, **13**, 492–506.

Jones, P. D., 1994: Hemispheric Surface Air Temperature Variations: A Reanalysis and an Update to 1993. *Journal of Climate*, **7**(11), 1794–1802.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, and D. Joseph, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*, **77**(3), 437–471.

Kelly, P. M., 1979: An Arctic Sea Ice Data Set, 1901–1956. Glaciological Data Report GD-5, World Data Center A for Glaciology (Snow and Ice).

Parker, D. E., and C. K. Folland, 1991: Worldwide Surface Temperature Trends Since the Mid-19th Century. In Schlesinger, M. E., editor, *Greenhouse Gas Induced Climatic Change: A Critical Appraisal of Simulations and Observations.*, 173–193. Elsevier Science Publishers B.V., Amsterdam.

——, ——, A. Bevan, M. N. Ward, M. Jackson, and K. Maskell, 1995a: Marine surface data for analysis of climatic fluctuations on interannual to century timescales. In Martinson, D. G., K. Bryan, M. Ghil, M. M. Hall, T. R. Karl, E. S. Sarachik, S. Sorooshian, and L. D. Talley, editors, *Natural Climate Variability on Decade-to-Century Time Scales*, 241–250. National Academy Press, Washington, D.C. colour figs. on pp222–228.

——, ——, and M. Jackson, 1995b: Marine surface temperature: observed variations and data requirements. *Climatic Change*, **31**, 559–600.

——, M. Jackson, and E. B. Horton, 1995c: The GISST2.2 sea surface temperature and sea-ice climatology. CRTN 63, Hadley Centre for Climate Prediction and Research, Meteorological Office, London Road, Bracknell, Berkshire, RG12 2SY.

Pickard, G. L., and W. J. Emery, 1982: *Descriptive Physical Oceanography: An Introduction.* Pergamon Press.

Potts, J. M., C. K. Folland, I. T. Jolliffe, and D. Sexton, 1996: Revised "LEPS" scores for assessing climate model simulations and long-range forecasts. *Journal of Climate*, **9**(1), 34–53.

Rayner, N. A., C. K. Folland, D. E. Parker, and E. B. Horton, 1995a. A new Global sea-Ice and Sea Surface Temperature (GISST) data set for 1903–1994 for forcing climate models. Hadley Centre Internal Note 69.

——, M. N. Ward, D. E. Parker, and C. K. Folland, 1995b: Using EOF Analysis to create a new GISST Data Set for Forcing GCMs. In Folland, C. K., and D. P. Rowell, editors, *Workshop on Simulations of the Climate of the Twentieth Century using GISST, 28–30 November 1994, Hadley Centre, Bracknell, UK.*, 52–53, Meteorological Office, London Road, Bracknell, Berkshire, RG12 2SY. Hadley Centre for Climate Prediction and Research. CRTN 56.

Reynolds, R. W., and T. M. Smith, 1994: Improved Global Sea Surface Temperture Analyses Using Optimum Interpolation. *Journal of Climate*, **7**(6), 929–948.

——, 1988: A Real-Time Global Sea Surface Temperature Analysis. *Journal of Climate*, **1**, 75–86.

Rowell, D. P., C. K. Folland, A. C. Renshaw, N. A. Rayner, and J. R. Davies, 1996: Simulated and observed seasonal variability and the influence of ocean boundary forcing. Hadley Centre Internal Note 73, Hadley Centre for Climate Prediction and Research, Bracknell, U.K. Final report to CEC on Medium Term Climate Variability; Contract $EV5V-CT92-0121$.

Smith, T. M., R. W. Reynolds, R. E. Livezey, and D. C. Stokes, 1996: Reconstruction of Historical Sea Surface Temperatures using Empirical Orthogonal Functions. *Journal of Climate*, **9**(6), 1403–1420.

Walsh, J. E., 1978: A Data Set on Northern Hemisphere Sea Ice Extent, 1953–76. Glaciological Data Report GD-2, World Data Center-A for Glaciology (Snow and Ice).

——, 1995: A sea ice database. In Folland, C. K., and D. P. Rowell, editors, *Workshop on Simulations of the Climate of the Twentieth Century using GISST, 28–30 November 1994, Hadley Centre, Bracknell, UK.*, 54–55, Meteorological Office, London Road, Bracknell, Berkshire, RG12 2SY. Hadley Centre for Climate Prediction and Research. CRTN 56.

Ward, M. N., 1989. On the estimation of EOF time coefficients when data are missing. Meteorological Office Synoptic Climatology Branch Discussion Note 114.

Woodruff, S. D., R. J. Slutz, R. L. Jenne, and P. M. Steurer, 1987: A Comprehensive Ocean-Atmosphere Data Set. *Bulletin of the American Meteorological Society*, **68**(10), 1239–1250.

Figure Captions

Figure 1. Derivation of an ice concentration gradient for part of the Antarctic (30°–15°W) in November 1980.
(a) shows the original GISST1.1 ice-edge and
(b) shows the GISST2.2 ice-edge derived therefrom.
Contours are tenths of ice concentration in each 1° box.

Figure 2. Part of the Coast of Greenland (40°W–0°) July 1987.
(a) shows the Walsh observed sea-ice concentration in tenths and (b) shows the ice-zone SST (°C) derived therefrom using June–Aug. regression relations between ice conc. and *in situ* SST for overlapping longitude bands.
Contours are at -1.79, 0, 1.5, 2, 3 and 4°C for SST and 0, 2, 4, 6, 8 and 10 tenths for sea-ice.

Figure 3. First EOF of low-pass filtered equal area GISST1.1 anomalies for 1901–1990 (a) spatial weights and (b) time series.

Figure 4. Leading area-weighted 2° resolution covariance EOFs of de-trended filled Atlantic MOHSST6. Spatial patterns and time series of (a) EOF 1, (b) EOF 2, (c) EOF 3 and (d) EOF 4.

Figure 5. As Figure 4, but for the Indian Ocean.

Figure 6. As Figure 4, but for the Pacific.

Figure 7. As Figure 4, but for the Mediterranean/Black Sea region.

Figure 8. Natural log of the first 50 eigenvalues of de-trended filled Pacific MOHSST6 anomalies. Marked modes were used in the test reconstructions.

Figure 9. Variance of seasonal 2° fields for 1982–1993 of (a) OI SST anomalies and anomalies reconstructed from the global trend EOF and (b) 4, (c) 21 and (d) 30 Pacific EOFs.

Figure 10. The La Niña event of May 1988 as represented in the (a) GISST1.1 and (b) GISST2.2 data sets. Anomalies (°C) are wrt. the 1961–90 GISST2.2 climatology.

Figure 11. The El Niño event of January 1983 as represented in the (a) GISST1.1 and (b) GISST2.2 data sets. Anomalies (°C) are wrt. the 1961–90 GISST2.2 climatology.

Figure 12. SST-ice concentration regression relations (dash-dotted lines) for each season (DJF, MAM, JJA, SON) in 31° longitude bands centred on (a) 70.5°W (Great Lakes excluded), (b) 30.5°W and (c) 129.5°E. Mean SSTs (°C) in each sea-ice concentration class are shown by the solid lines, concentration is in tenths.

Figure 13. SST-ice concentration regression relations (dash-dotted lines) for each season in (a) the whole Northern Hemisphere (Great Lakes excluded), (b) the whole Southern Hemisphere and (c) the freshwater Great Lakes. Mean SSTs (°C) in each sea-ice concentration class are shown by the solid lines, concentration is in tenths.

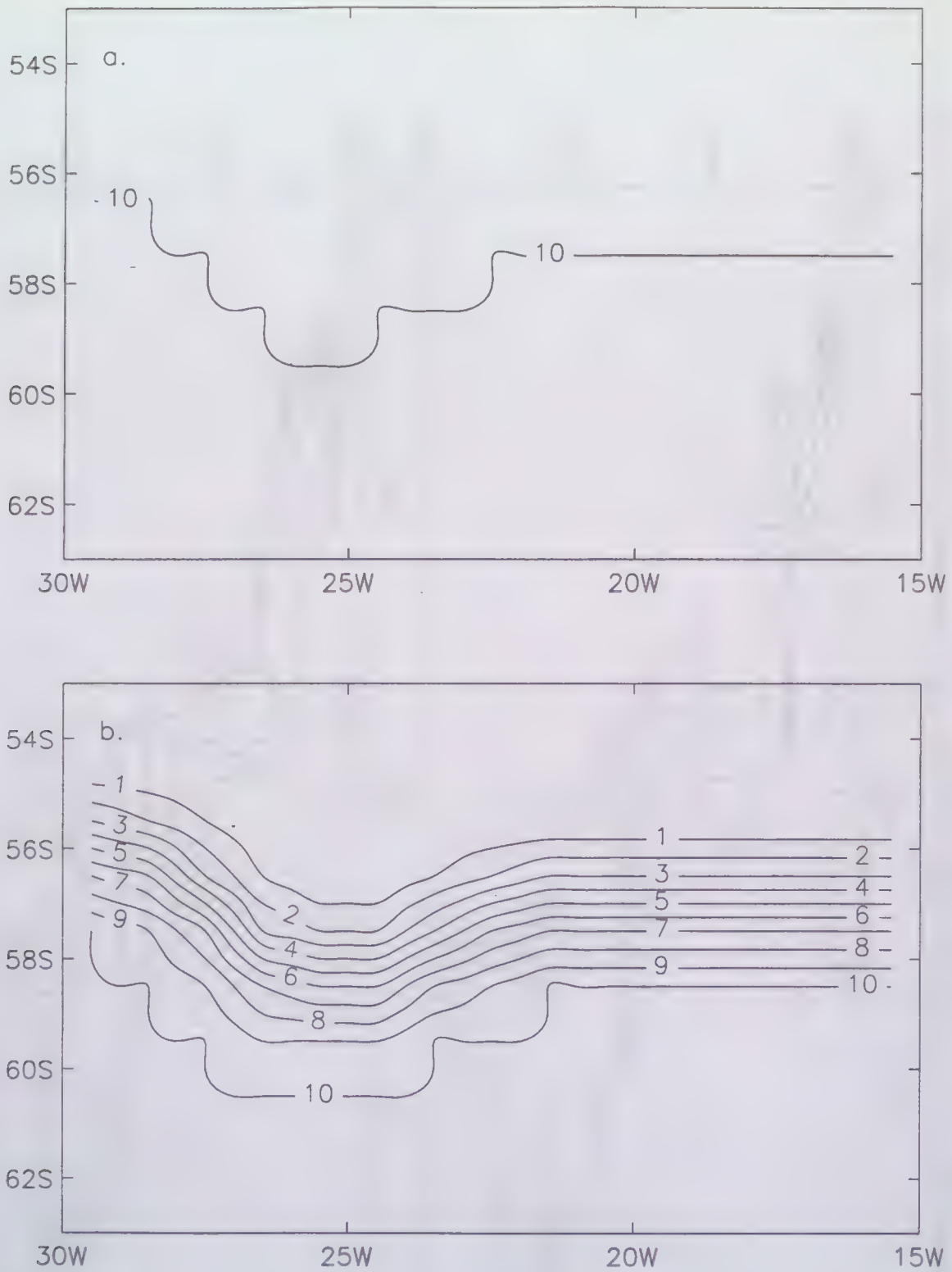

Figure 1. Derivation of an ice concentration gradient for part of the Antarctic (30°–15°W) in November 1980.
(a) shows the original GISST1.1 ice–edge and
(b) shows the GISST2.2 ice–edge derived therefrom.
Contours are tenths of ice concentration in each 1° box.

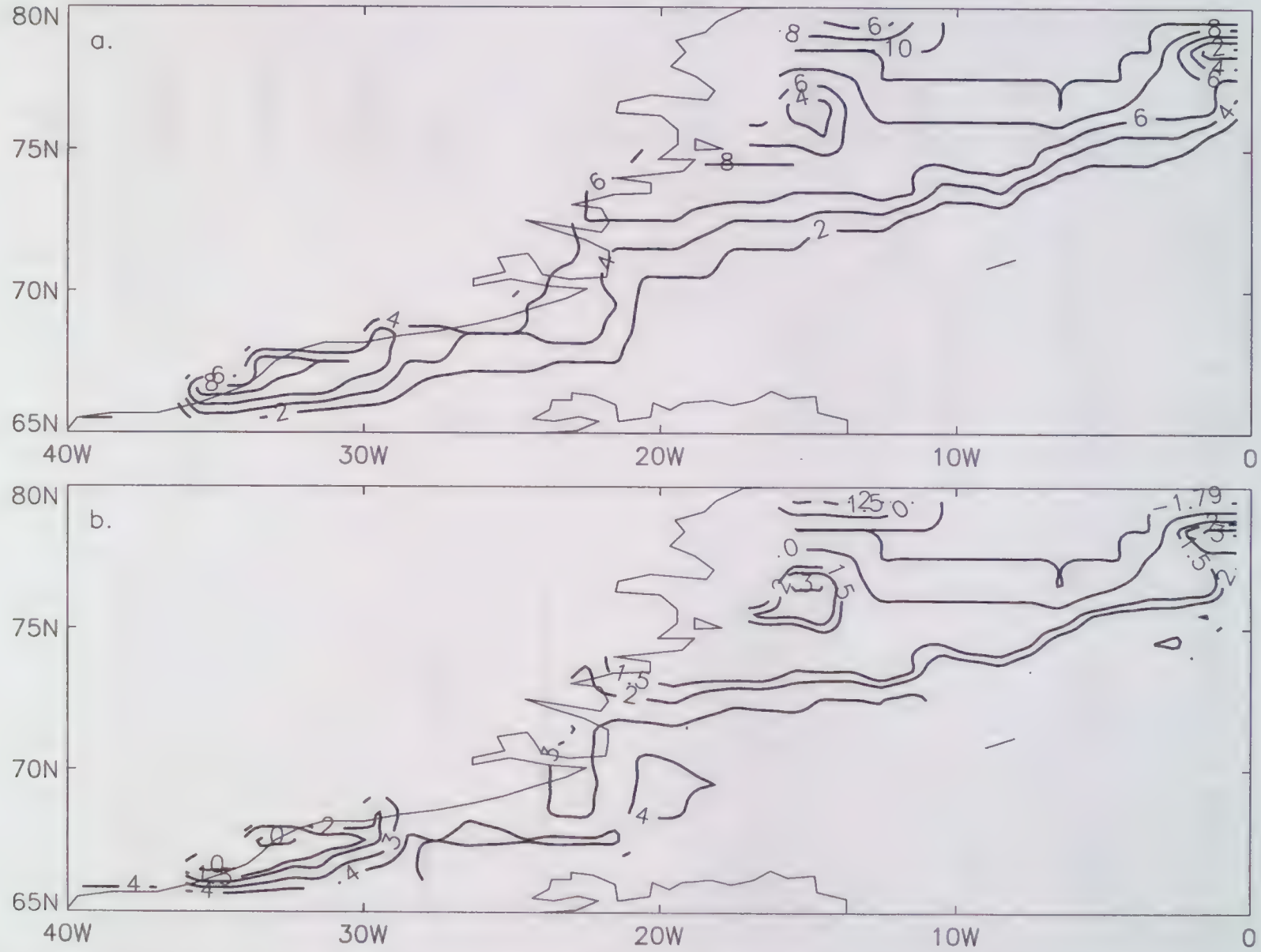

Figure 2. Part of the Coast of Greenland (40°W–0°) July 1987. (a) shows the Walsh observed sea–ice concentration in tenths and (b) shows the ice–zone SST (°C) derived therefrom using June–Aug. regression relations between ice conc. and in situ SST for overlapping longitude bands.
Contours are at −1.79, 0, 1.5, 2, 3 and 4 °C for SST and 0, 2, 4, 6, 8 and 10 tenths for sea–ice.

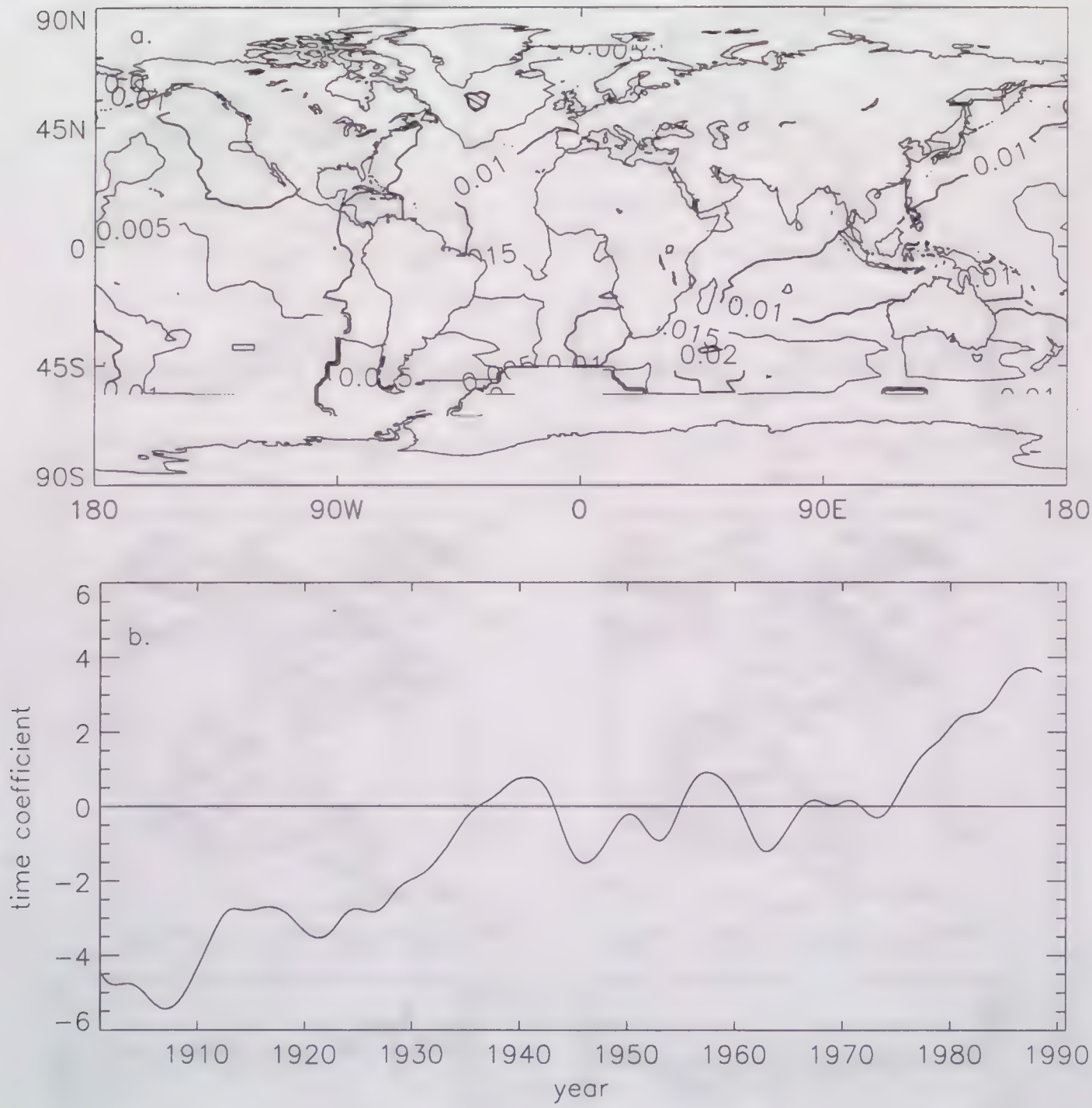

Figure 3. First EOF of low-pass filtered equal area GISST1.1 anomalies for 1901-1990 (a) spatial weights (b) time series.

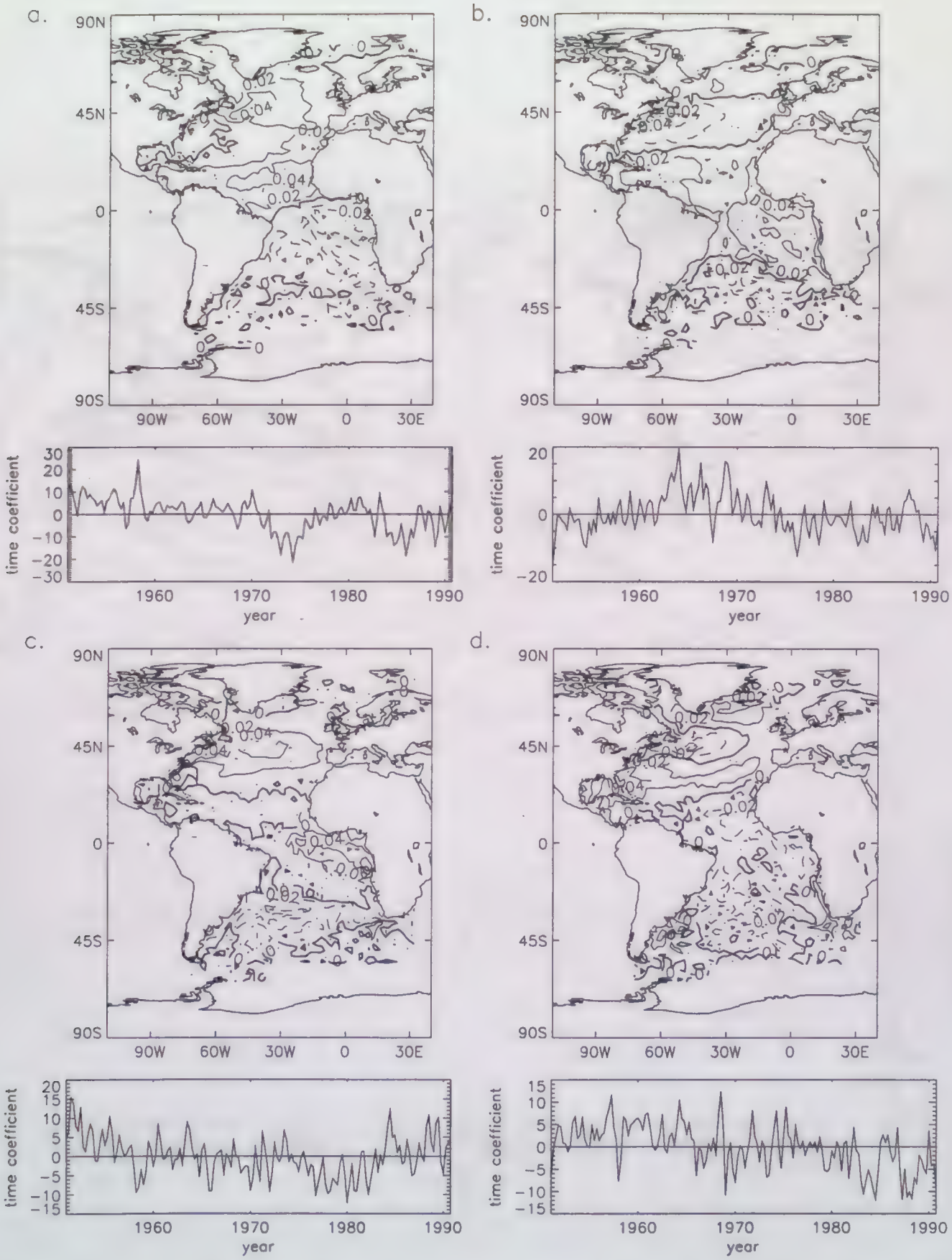

Figure 4. Leading area-weighted 2° resolution covariance EOFs of de-trended filled Atlantic MOHSST6. Spatial patterns and time series of EOFs (a) 1, (b) 2, (c) 3 and (d) 4.

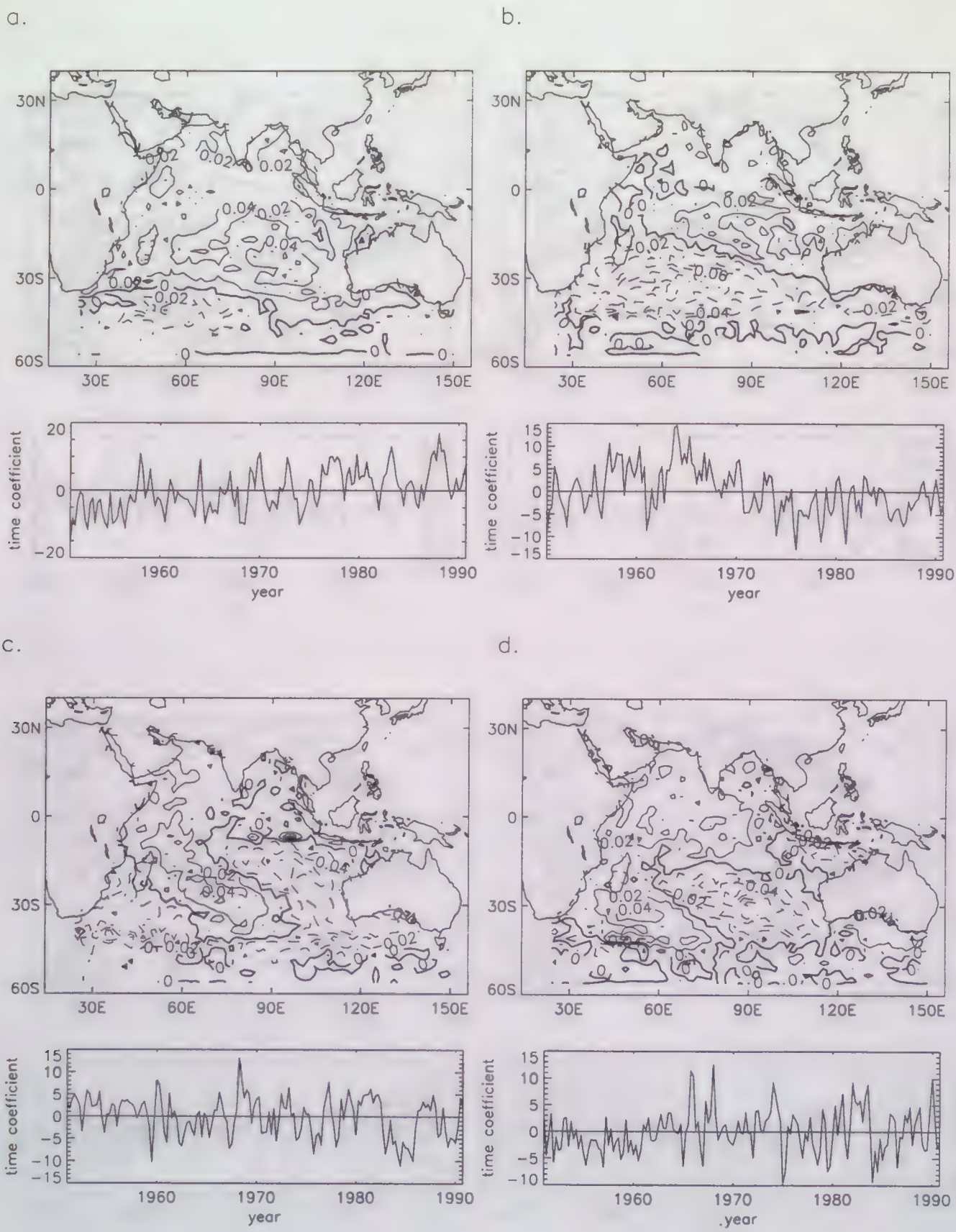

Figure 5. As Figure 4, but for the Indian Ocean.

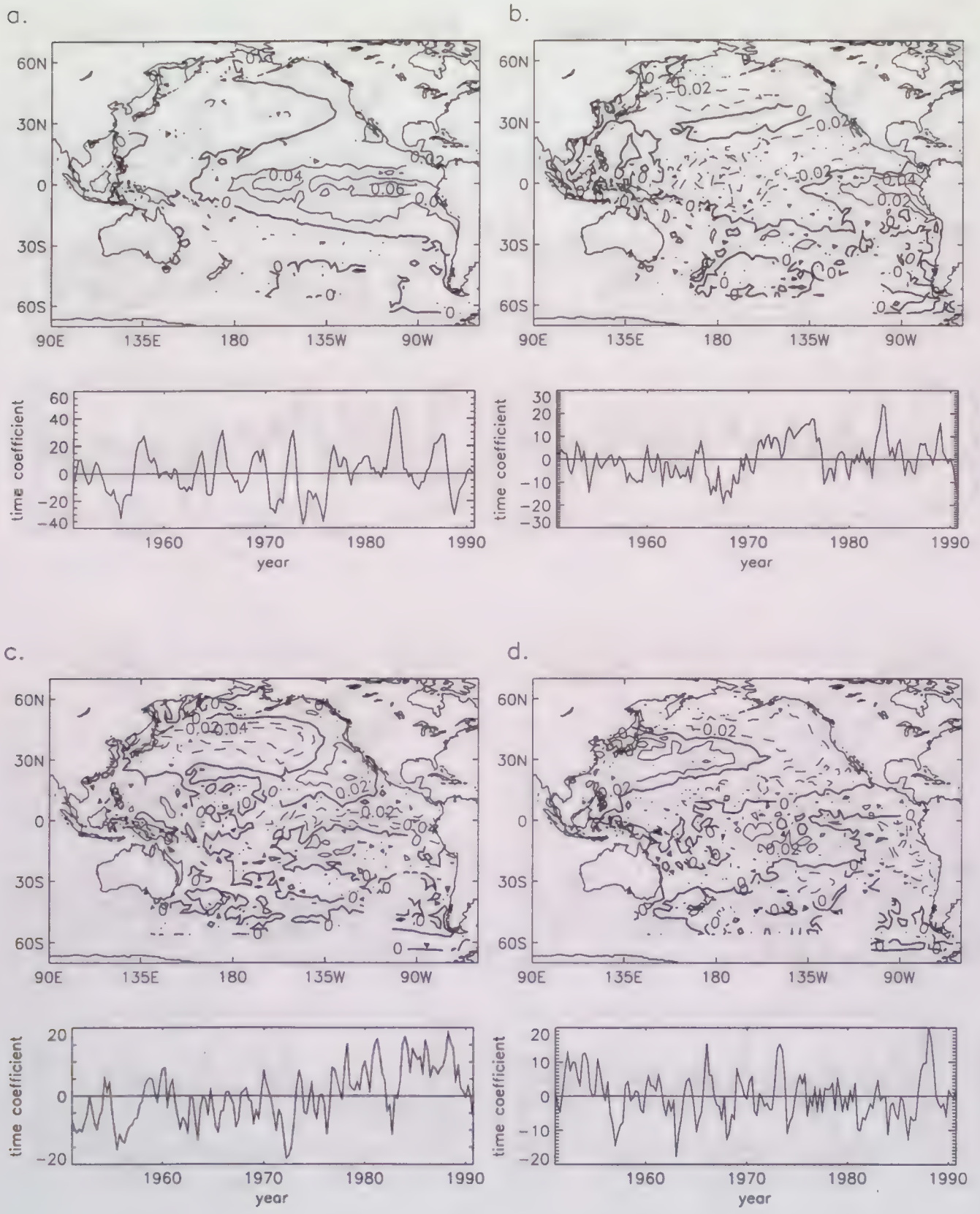

Figure 6. As Figure 4, but for the Pacific.

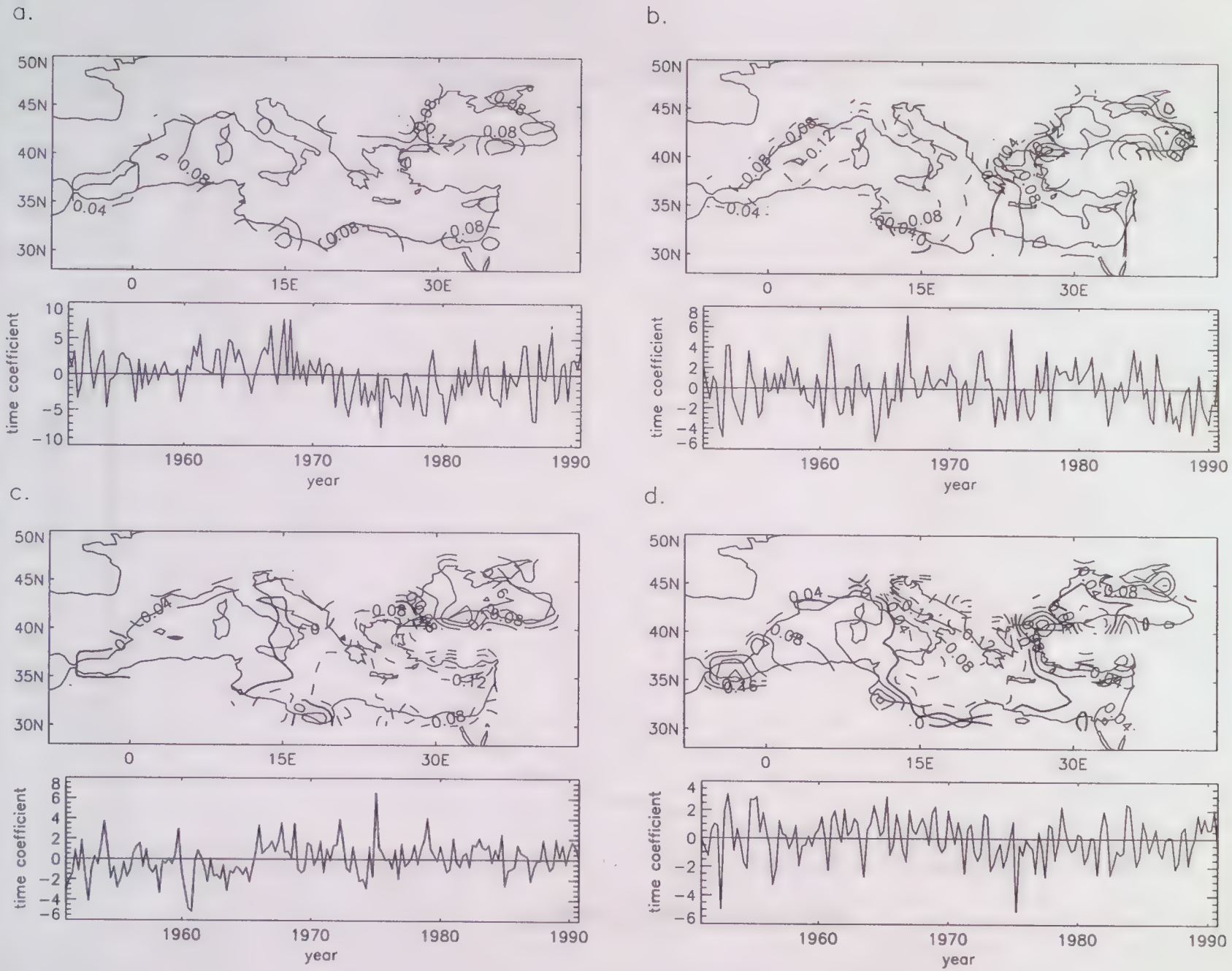

Figure 7. As Figure 4, but for the Mediterranean/Black Sea region.

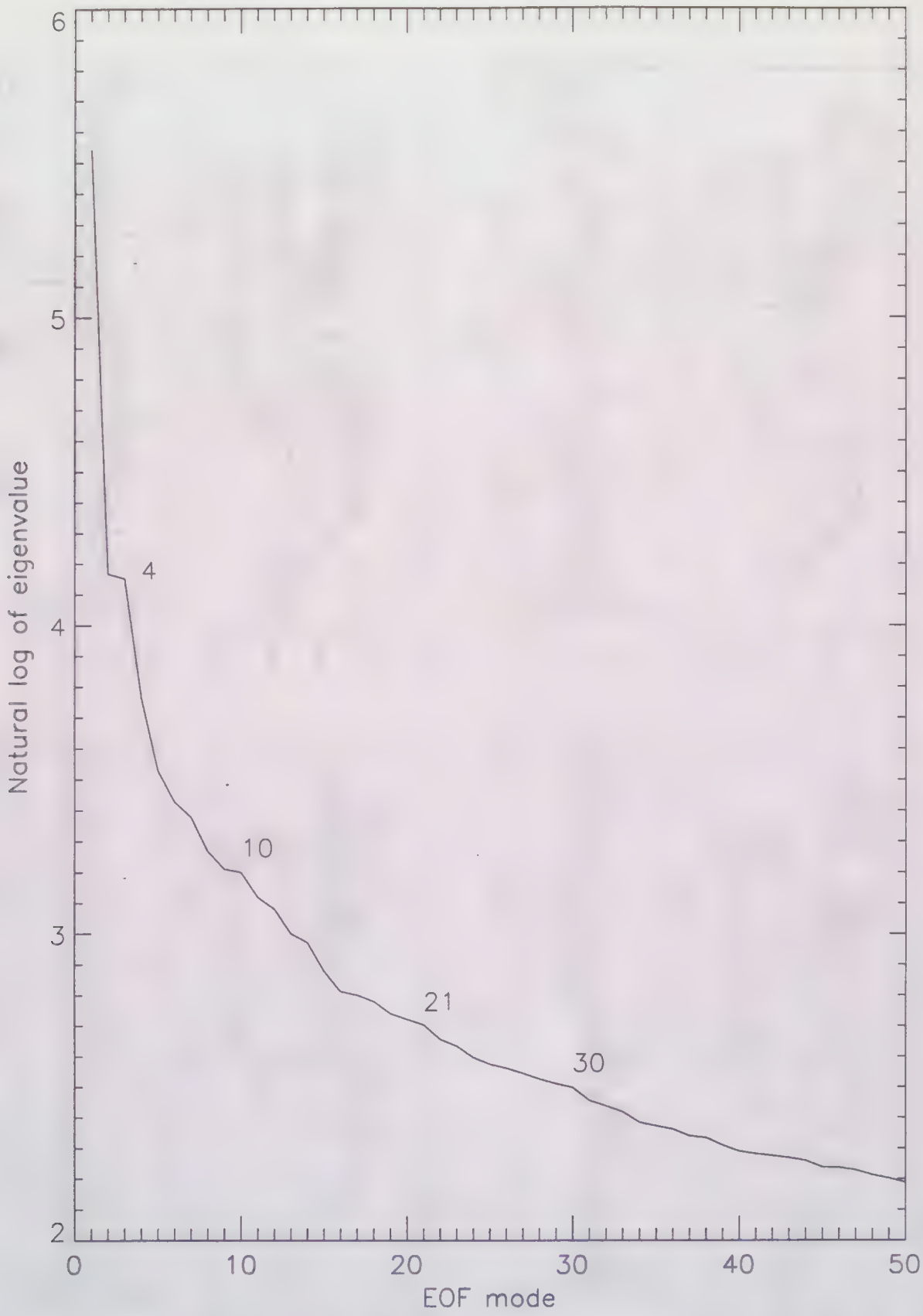

Figure 8. Natural log of the first 50 eigenvalues of de-trended filled Pacific MOHSST6 anomalies. Marked modes were used in the test reconstructions.

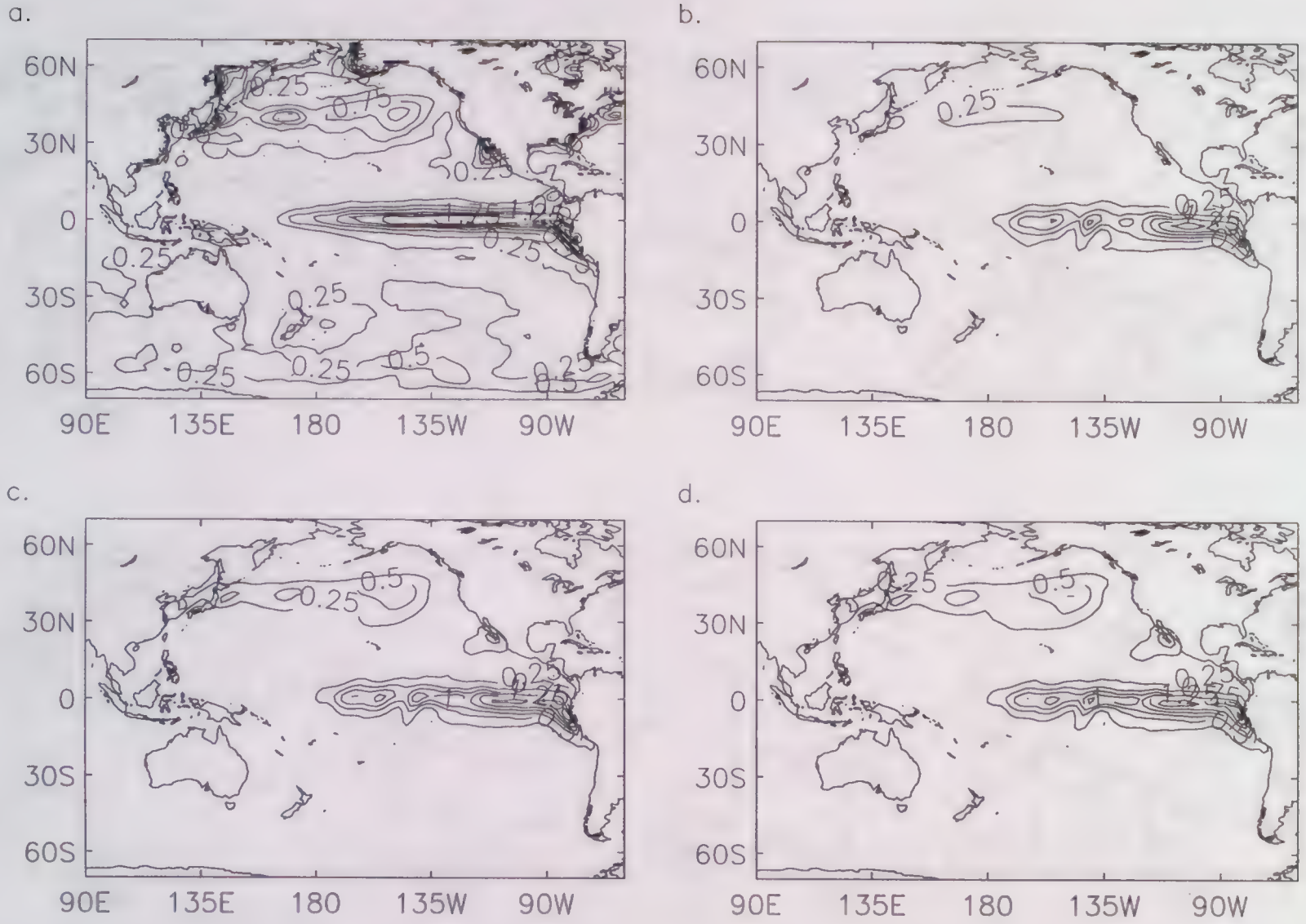

Figure 9. Variance of seasonal 2° fields for 1982–1993 of (a) OI SST anomalies and anomalies reconstructed from the global trend EOF and (b) 4, (c) 21 and (d) 30 Pacific EOFs.

a.

b.

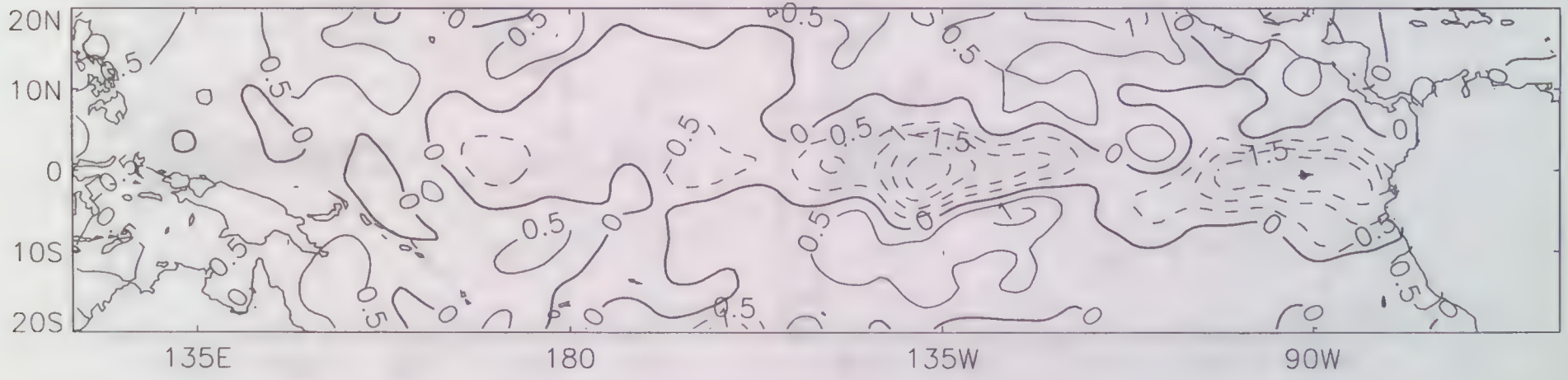

Figure 10. The La Nina event of May 1988 as represented in the (a) GISST1.1 and (b) GISST2.2 data sets. Anomalies (°C) are wrt. the 1961–90 GISST2.2 climatology.

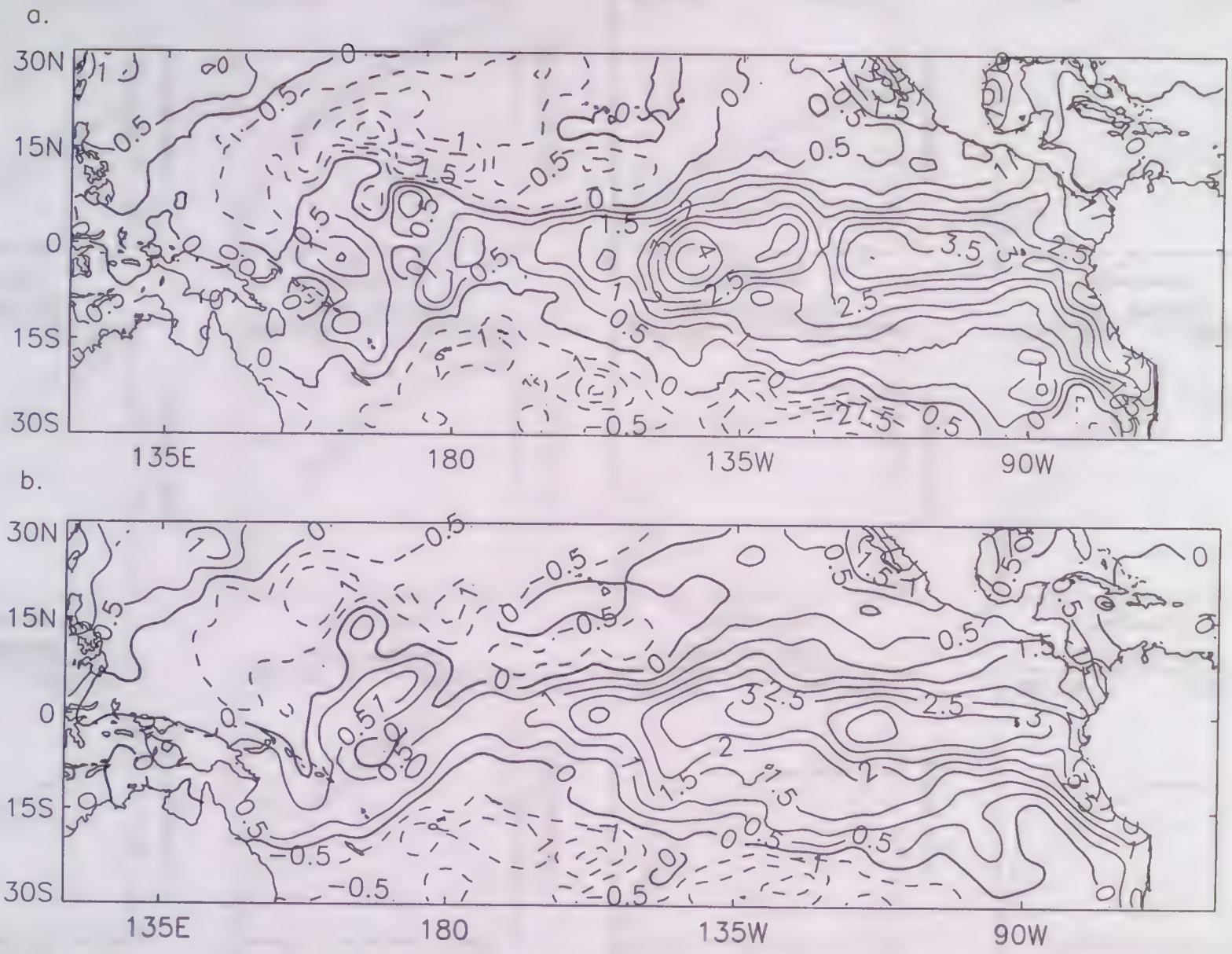

Figure 11. The El Nino event of January 1983 as represented in the (a) GISST1.1 and (b) GISST2.2 data sets. Anomalies (°C) are wrt. the 1961–90 GISST2.2 climatology.

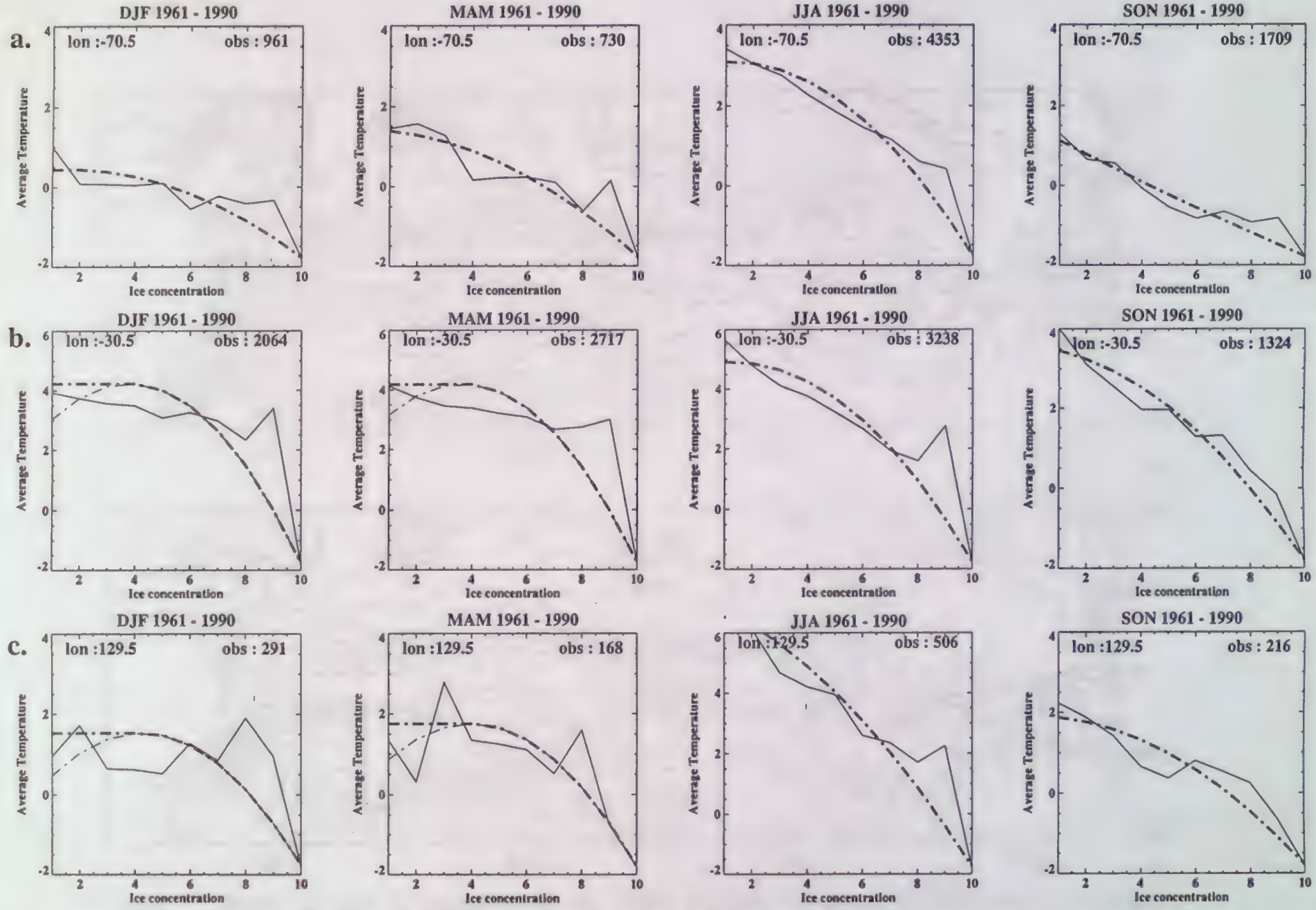

Figure 12. SST-ice concentration regression relations (dash-dotted lines) for each season (DJF, MAM, JJA, SON) in 31° longitude bands centred on (a) 70.5°W (Great Lakes excluded) (b) 30.5°W and (c) 129.5°E. Mean S STs (°C) in each sea-ice concentration class are shown by the solid lines, concentration is in tenths.

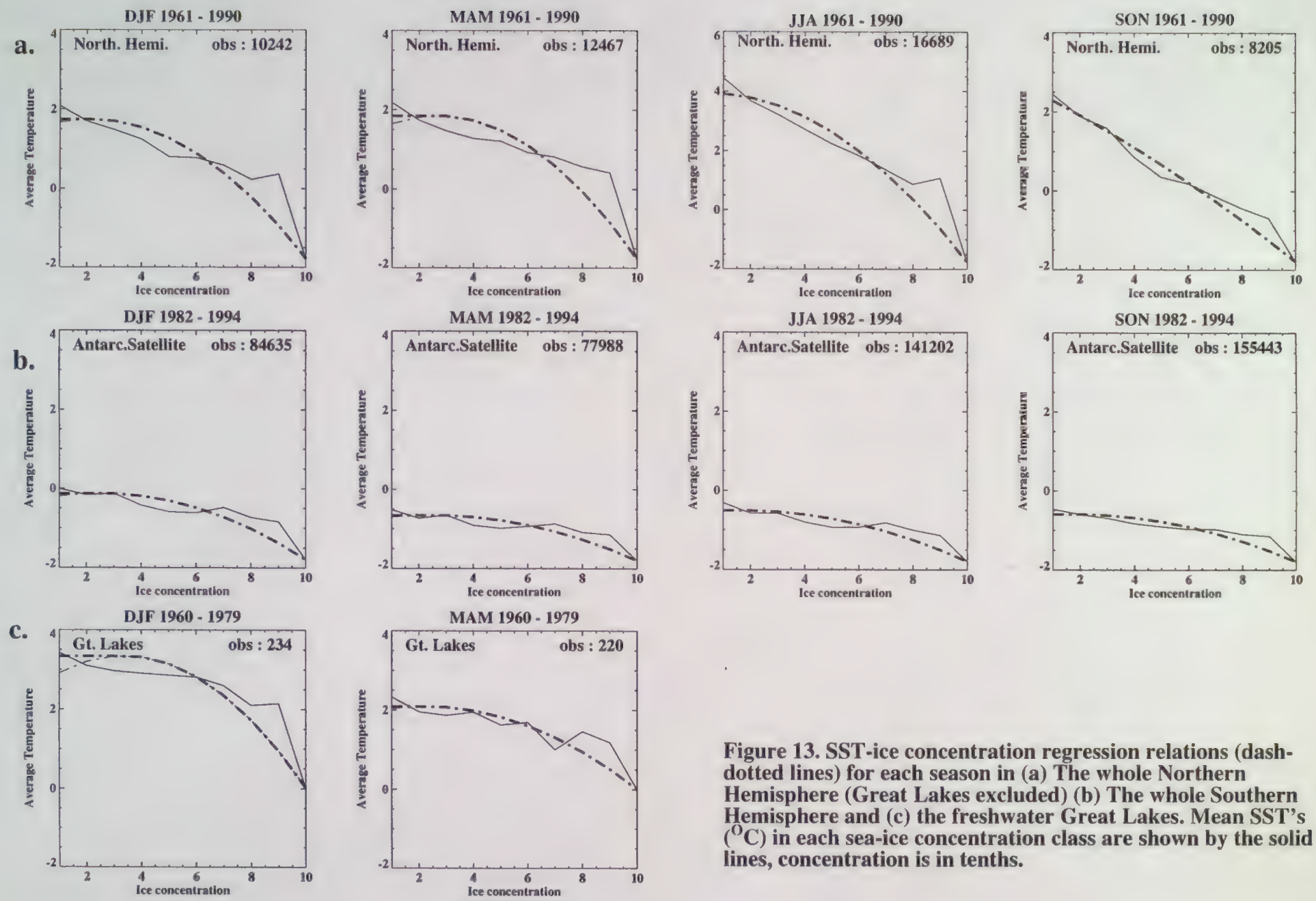

Figure 13. SST-ice concentration regression relations (dash-dotted lines) for each season in (a) The whole Northern Hemisphere (Great Lakes excluded) (b) The whole Southern Hemisphere and (c) the freshwater Great Lakes. Mean SST's (^{O}C) in each sea-ice concentration class are shown by the solid lines, concentration is in tenths.

CLIMATE RESEARCH TECHNICAL NOTES

CRTN 1	Oct 1990	Estimates of the sensitivity of climate to vegetation changes using the Penman-Monteith equation. P R Rowntree
CRTN 2	Oct 1990	An ocean general circulation model of the Indian Ocean for hindcasting studies. D J Carrington
CRTN 3	Oct 1990	Simulation of the tropical diurnal cycle in a climate model. D P Rowell
CRTN 4	Oct 1990	Low frequency variability of the oceans. C K Folland, A Colman, D E Parker and A Bevan
CRTN 5	Dec 1990	A comparison of 11-level General Circulation Model Simulations with observations in the East Sahel. K Maskell
CRTN 6	Dec 1990	Climate Change Prediction. J F B Mitchell and Qing-cun Zeng
CRTN 7	Jan 1991	Deforestation of Amazonia - modelling the effects of albedo change. M F Mylne and P R Rowntree
CRTN 8	Jan 1991	The role of observations in climate prediction and research. D J Carson
CRTN 9	Mar 1991	The greenhouse effect and its likely consequences for climate change. D J Carson
CRTN 10	Apr 1991	Use of wind stresses from operational N.W.P. models to force an O.G.C.M. of the Indian Ocean. D J Carrington
CRTN 11	Jun 1991	A new daily Central England Temperature series, 1772-1991. D E Parker, T P Legg and C K Folland
CRTN 12	Jul 1991	Causes and predictability of Sahel rainfall variability. D P Rowell, C K Folland, K Maskell, J A Owen, M N Ward

CLIMATE RESEARCH TECHNICAL NOTES

CRTN 13	Jul 1991	Modelling changes in climate due to enhanced CO2, the role of atmospheric dynamics, cloud and moisture. C A Senior, J F B Mitchell, H Le Treut and Z-X Li
CRTN 14	Sep 1991	Sea temperature bucket models used to correct historical SST data in the Meteorological Office. C K Folland
CRTN 15	Aug 1991	Modelling climate change, and some potential effects on agriculture in the U.K. P R Rowntree, B A Callander and J Cochrane
CRTN 16	Aug 1991	The Boreal Forests and Climate G Thomas and P R Rowntree
CRTN 17	Aug 1991	Development of a Stratosphere-Troposphere Data Assimilation System. R Swinbank.
CRTN 18	Sep 1991	A study of asynchronous coupling using a simple climate model. M K Davey.
CRTN 19	Sep 1991	The Oceanic Carbon Cycle. N K Taylor
CRTN 20	Nov 1991	Worldwide ocean-atmosphere surface fields in Sahel wet and dry years using provisionally corrected surface wind data. M N Ward
CRTN 21	Dec 1991	Coupled tropical ocean global atmosphere models at the UKMO. M K Davey, C Gordon, S Ineson and S Lawrence
CRTN 22	Dec 1991	Empirical parameterisation of tropical ocean-atmosphere coupling. M Allen, M K Davey, D L T Anderson and P D Killworth
CRTN 23	Feb 1992	The temporal evolution of Equatorial Currents in the Indian Ocean D L T Anderson, D J Carrington, R A Corry and C Gordon
CRTN 24	Feb 1992	Stratospheric analyses provided by the U.K. Meteorological Office M J Bailey, A O'Neill and V D Pope

CLIMATE RESEARCH TECHNICAL NOTES

CRTN 25	Feb 1992	Modelling interannual variability in the Indian Ocean using momentum fluxes from the UKMO and ECMWF operational weather analyses. D L T Anderson and D J Carrington
CRTN 26	Mar 1992	A GCM simulation of the impact of Amazonian deforestation on climate using an improved canopy representation J Lean and P R Rowntree
CRTN 27	Mar 1992	The parameterization of rainfall interception in GCMs A J Dolman and D Gregory
CRTN 28	Jun 1992	Development of worldwide marine data eigenvectors since 1985 A W Colman
CRTN 29	Jun 1992	A tropical ocean model with reduced physics M K Davey
CRTN 30	Jun 1992	International Temperature Workshop, Boulder, Colorado, USA, 16 January 1992 Edited by D E Parker
CRTN 31	Jul 1992	Simulation of clear-sky outgoing longwave radiation over the oceans using operational analyses A Slingo and M J Webb
CRTN 32	Sep 1992	A prediction of the transient response of climate J M Murphy
CRTN 33	Nov 1992	LEPS scores for assessing climate model simulations and long-range forecasts. C K Folland
CRTN 34	Jan 1993	Stratospheric Data Assimilation System Guide Editer: R Swinbank
CRTN 35	Mar 1993	A Stratosphere-Troposphere Data Assimilation System R Swinbank and A O'Neill
CRTN 36	Mar 1993	Validation of hydrological schemes for climate models against catchment data. P R Rowntree and J Lean

CLIMATE RESEARCH TECHNICAL NOTES

CRTN 37	Apr 1993	Modelling of palaeoclimates: Examples from the recent past J F B Mitchell
CRTN 38	May 1993	A simulation of seasonality in ENSO forecast skill M K Davey, D L T Anderson and S Lawrence
CRTN 39	Jul 1993	ENSO Prediction experiments using a simple ocean-atmosphere model D-H Wu, D L T Anderson and M K Davey
CRTN 40	Sep 1993	Diagnosis of dynamic sea-surface and sea-level changes from the Cox ocean model J M Gregory
CRTN 41	Sep 1993	Seasonal variations of the clear-sky greenhouse effect: the role of changes in atmospheric temperatures and humidities M J Webb, A Slingo and G L Stephens
CRTN 42	Nov 1993	Sea-level changes under increasing atmospheric CO_2 in a transient coupled ocean-atmosphere GCM experiment. J M Gregory
CRTN 43	Dec 1993	Global and regional patterns of climate change: recent predictions for the Arctic P R Rowntree
CRTN 44	Feb 1994	The effect of changing horizontal diffusion in the atmospheric version of the unified climate model C D Hall and R A Stratton
CRTN 45	Mar 1994	Simulation of El-Niño/Southern Oscillation like variability in a global AOGCM and its response to CO_2 increase. S F B Tett
CRTN 46	Apr 1994	Global data required for monitoring climate change. D E Parker and C K Folland
CRTN 47	May 1994	Seasonal uptake of anthropogenic CO_2 in an ocean general circulation model. N K Taylor

CLIMATE RESEARCH TECHNICAL NOTES

CRTN 48	Jun 1994	A tropic-wide oscillation of boreal summer rainfall and patterns of sea-surface temperature. M N Ward, K Maskell, C K Folland, D P Rowell and R Washington
CRTN 49	Jun 1994	Simulation of the tropical Pacific using a simplified ocean model M A Balmaseda, D L T Anderson and M K Davey
CRTN 50	Sep 1994	Simulation of global mean temperature using a box-diffusion climate model P R Rowntree
CRTN 51	Sep 1994	Seasonal dependence of ENSO prediction skill M A Balmaseda, M K Davey and D L T Anderson
CRTN 52	Nov 1994	Simulation of the Indian Monsoon and Tropical Intraseasonal Variability by a General Circulation Model P M Inness and D Gregory
CRTN 53	Mar 1995	A comparison of modelled surface fluxes with climatological estimates. C Gordon and D K Wright
CRTN 54	Mar 1995	The representation of moist convection in atmospheric models. D Gregory
CRTN 55	Mar 1995	A consistent treatment of the evaporation of rain and snow for use in large-scale models D Gregory
CRTN 56	Apr 1995	Workshop on simulations of the Climate of the Twentieth Century using GISST, 28-30 November 1994, Hadley Centre, Bracknell, UK Edited by C K Folland and D P Rowell
CRTN 57	May 1995	Enhanced shortwave cloud radiative forcing due to anthropogenic aerosols. S E Swartz and A Slingo
CRTN 58	May 1995	The simulation of the tropical oceans in models of different horizontal resolution. C Gordon, D K Wright and C M Roberts
CRTN 59	Sep 1995	The water budget of middle latitude continental regions - a modelling and observational study P R Rowntree
CRTN 60	Sep 1995	Intraseasonal variability of the Indian summer monsoon simulated by the Hadley Centre climate model K Ashok, D Gregory and P M Inness

CLIMATE RESEARCH TECHNICAL NOTES

CRTN 61	Sep 1995	Climate simulations with the Unified Model: AMIP runs C D Hall, R A Stratton and M L Gallani
CRTN 62	Nov 1995	Validation of surface parameters over the oceans in climate simulations with the unified model. C D Hall
CRTN 63	Dec 1995	The GISST2.2 sea surface temperature and sea-ice climatology D E Parker, M Jackson and E B Horton
CRTN 64	Dec 1995	Parametrization of momentum transport by convection. II: Tests in single column and general circulation models. D Gregory, R Kershaw and P M Inness
CRTN 65	Jan 1996	Understanding the sensitivity of a GCM simulation of Amazonian deforestation to the specification of vegetation and soil characteristics. J Lean and P R Rowntree
CRTN 66	Jan 1996	On the efficient calculation of infra-red fluxes and cooling rates using the two-stream equations. J M Edwards
CRTN 67	Jan 1996	The second Hadley Centre coupled ocean-atmosphere GCM: Model description, spinup and validation. TC Johns,RE Carnell,JF Crossley,JM Gregory, JFB Mitchell,CA Senior,SFB Tett and RA Wood
CRTN 68	Feb 1996	The sensitivity of climate simulations to the specification of mixed phase clouds D Gregory and D Morris
CRTN 69	Apr 1996	Using an ensemble of Multi-Decadal GCM simulations ro assess potential seasonal predictability D P Rowell
CRTN 70	May 1996	North Atlantic and European Seasonal Predictability using an Ensemble of Multi-Decadal AGCM Simulations J R Davies, D P Rowell and C K Folland
CRTN 71	Jun 1996	A description of the Second Hadley Centre Coupled Model (HADCM2) T C Johns
CRTN 72	Jul 1996	Palaeoclimate Modelling Intercomparison: UKMO GCM simulations for 6kBP and 21kBP C D Hewitt and J F B Mitchell

CLIMATE RESEARCH TECHNICAL NOTES

CRTN 73	Aug 1996	ENSO responses and low frequency weather variability in the North Pacific/American sector 1949-93 A C Renshaw, D P Rowell and C K Folland
CRTN 74	Sep 1996	Version 2.2 of the Global sea-Ice and Sea Surface Temperature data set, 1903-1994 N A Rayner, E B Horton, D E Parker, C K Folland and R B Hackett